Understanding Biological Timing by Modelling Simple Circadian Clocks

Oxana Sorokina

Doctor of Philosophy

Institute of Molecular Plant Sciences

The University of Edinburgh

November 2009

Table of Contents

Declarations

This Thesis contains work based on the following published papers.

1) A switchable light-input, light-output system modelled and constructed in yeast, **Oxana Sorokina**. Anita Kapus, Kata Terecskei, Laura E Dixon, Laszlo Kozma-Bognar, Ferenc Nagy, Andrew J Millar
Journal of Biological Engineering 2009, **3**:15 (17 September 2009)

I hereby declare that this Thesis is my own work except where explicitly stated.

No part of this thesis has been submitted for a professional qualification or a degree at the University of Edinburgh or any other university.

Oxana Sorokina

November, 2009

Acknowledgements

I would like first to thank Professor Andrew J. Millar his excellent supervision of my project. Boundless discussion on scientific work and rationale to carry out the project were wonderful. Three years of working with him were a great experience for me.

I would also like to deeply acknowledge my husband, Anatoly Sorokin, for his patience and carrying out the entire family job when needed, and for his high expertise in Systems Biology and fruitful discussions that successfully contributed to this wok.

I would like to say special thanks to Laszlo Kozma- Bognar for the extremely productive collaboration and keeping me cheerful irrespectively of the state of the experiment-model relationships.

I would love to give special thanks to Laura Dixon and Owain Bristow for proofreading part of my work.

Abstract

Circadian clocks involve complex gene regulatory network that influences the overall organism physiology through the rhythmic expression of the genes responsible for key metabolic pathways. It receives external cues in the form of light and temperature inputs to synchronize the internal cycle with environmental 24h day-night cycle.

We are using the Synthetic Biology approach to uncover the fundamental principles underlying the evolution and principle operation of the circadian clock. This approach allows the testing and exploration of potential applications that are not limited by natural, evolved systems. To address this issue we start from generation of hypothetical principles with the aid of mathematical modelling, following by testing artificial biological networks based on these principles.

We are developing synthetic gene networks in yeast to reproduce the experimental data and modelling results obtained for evolved circadian clocks. We show that a one-loop circuit could exhibit entrained, periodic oscillations. The test circuit is based on the Tet-On gene expression system and includes one artificial light input, based on the light-switchable phytochrome-based promoter system. A mathematical model of this circuit was proposed to simulate, explain and predict the possible behaviour of this network. Two reporter systems for the circuit were tested and modelled to provide an adequate readout.

We are also using the naturally reduced model organism *Ostreococcus tauri* to propose a model for the circadian regulation of a crucial metabolic pathway for starch synthesis and degradation. Based on literature data we consider the predominance of the genetic regulation in the unicellular alga compared to the higher plants. Proceeding from this assumption we developed an approach which allows making valid prediction of circadian regulatory pattern for starch metabolism in *Ostreococcus* based on gene expression data.

Abbreviations

ADPG	ADP- glucose
AGPase	ADP-glucose Phosphorylase
BG	Branched Glucans
bHLH	Basic-helix-loop-helix
BMAL1	Brain and muscle Arnt-like protein-1 gene
CCA1	CIRCADIAN CLOCK ASSOCIATED 1 gene
CKII	Casein Kinase II
CLK	CLOCK gene
CoSet	correlation reaction subset
CRY	Cryptochrome gene
CYC	CYCLE gene
DBT	DOUBLE-TIME gene
dox	Doxycycline
ELF3	EARLY FLOWERING 3 gene
ELF4	EARLY FLOWERING 4 gene
EST	Expressed sequence tag
F6P	Fructose- 6 Phosphate
FBA	Flux Balance Analysis
FHL	FAR-RED ELONGATED HYPOCOTYL LIKE gene
FHY1	FAR-RED ELONGATED HYPOCOTYL 1 gene
FKF1	FLAVIN-BINDING, KELCH REPEAT, F BOX 1 ligase
FRL	Far Red Light
FRQ	FREQUENCY gene
FVA	Flux Variability Analysis
G1P	Glucose-1- phosphate
G6P	Glucose-6-phosphate
GAD	GAL4- transcriptional activator
GAP	glyceraldehyde-3- phosphate
GBD	GAL4- DNA-binding domain

GBSSI	granule- bound starch synthase
GI	GIGANTEA gene
GWD	Glucan , Water Dikinase
HB	Hopf bifurcation
HTT	High-throughput technologies
IEA	Inverse Eigenvalues Analysis
iGEM	international Genetically Engineered Machine competition
KEGG	Kyoto Encyclopedia of Genes and Genomes databases
LG	Linear Glucans
LHY	LATE ELONGATED HYPOCOTYL gene
LUC	luciferase gene
LUX	LUX ARRHYTHMO gene
MEX1	Arabidopsis maltose transporter
NLS	Nuclear Localisation Sequence
ODE	Ordinary Differential Equation
PAS	Per – Period circadian protein, Arnt-Ah receptor, Sim single-minded protein
PCA	principal component analysis
PDP1	Par- Domain Protein 1 gene
PER	PERIOD gene
PHYA	Phytochrome A gene
PHYB	Phytochrome B gene
PIF	PHYTOCHROME INTERACTION FACTOR gene
PIL	PHYTOCHROME INTERACTION FACTOR-LIKE gene
PRC	Phase Response curve
PRR	PSEUDO RESPONSE REGULATOR
PWD	Phosphoglucan , Water Dikinase
RFU	Relative Flux Units
RL	Red Light
RLU	Relative Light Units
rtTA	tetracycline reverse transcriptional activator
SAPs	Sequestrated Areas of Phytochrome

SAPs	Sequestrated areas of fhytochrome
SBE	starch branching enzyme
SBGN	Systems Biology Graphical Notation
SBML	Systems Biology Markap language
SBTOOLBOX2	Systems Biology Toolbox 2
SEX4	Starch Excess 4 gene
SN	saddle-node
SNIC	saddle-node on invariant circuits
SS	Starch Synthase
Starch Flux	Flux through 'starch_ex' exchange reaction
Tc	Tetracycline
TIM	TIMELESS gene
TOC1	TIMING OF CAB EXPRESSION gene
Trx	thioredoxin
tTa	tetracycline transcriptional activator
UAS	Upstream Activation Sequence
VRI	VRILLE gene
WC-1	White Collar-1 gene
WC-2	White Collar -2 gene
WCC	White Collar Complex
ZT	zeitgeber time

List of figures

List of tables

Chapter 1. Introduction

1.1. Molecular Systems Biology: From Data to Model

A major hypothesis of molecular systems biology is that the behaviour of complex systems can be described by taking into account properties of simple atomic units of the system and the interactions between them. The investigation of the properties of interactions within the system is much more difficult than the investigation of the properties of its parts, mainly because those interactions are dynamic and transient. Being an abstraction of reality, modelling allows one to focus on the main features and components of the system. This is why modelling has become such a popular and important part of the systems biology approach.

In biological sciences a model implies a description of the current view on the structure and principles of behaviour of the system in textual form or as a diagram. Such descriptions are informal, non rigorous, ambiguous and difficult to reuse. However, these models are very useful as an essence of biological knowledge. On the contrary, mathematical models are very formal, concise and unambiguous, but this strictness requires extra assumptions to fill gaps in biological knowledge.

Modern literature classifies modelling approaches as data driven or mechanistic, or as inductive or deductive, even as Baconian or Popperian based upon the direction of knowledge flow from details and experimental data in data-driven and inductive models or from basic laws in deductive or mechanistic models [1]. From the point of view of their explanatory ability, models can also be classified into the explanatory and descriptive.

An explanatory model is able to shed light on the design, basic principles and mechanism of system functioning, while a descriptive approach just gives the

ability to predict the behaviour of the system based upon knowledge of the previous state of the system.

Data-driven modelling

High-throughput technologies (HTT) allow the measurement of countless numbers of different components of the system simultaneously. Statistical modelling techniques like clustering, principal component analysis (PCA) and regression modelling [2] are known as data-driven modelling, because little or no prior knowledge of the system is used [3].

Qualitative modelling

Qualitative modelling describes the coarse state of the system and its change over discrete time intervals, which yields an understanding of some mechanisms driving the system behaviour; therefore, this class of techniques is more explonatory than data-driven modelling. Hypotheses, which could be tested with qualitative models generally include logical and causal relationships between events, stimuli and system responses [3].

Constraint-based modelling

As a result of intrinsic homeostasis, biological systems can be treated in many cases as being in a steady state. In this case the numbers of generated and degraded molecules are equal and the overall state of the system is almost constant in time. This constant state depends on the structure of the system, for example net rates of biochemical reactions and their stoichiometry, general thermodynamics laws and the intensity of environmental signals, for instance the concentration of glucose in the intracellular environment [4].

Constraint based analysis is able to identify the set of fluxes through the network, which are compatible with the desired steady state, to predict missing reactions, alternative or unused pathways, and to find the set steady state fluxes, which is accessible in response to environmental signals.

Though it requires much more prior knowledge about the system in comparison with the data-driven and qualitative modelling, it demonstrates, nevertheless, greater explanatory power.

Quantitative modelling: Ordinary Differential Equations (ODEs)

In spite of the fact that the steady state assumption is generally applicable for the whole system, the steady state itself is not of great importance for the investigation of signal transduction pathways, circadian rhythms or the cell cycle, when the dynamics of the system becomes the point of interest. In this case quantitative modelling techniques provide a detailed explanation of the mechanisms and allow numerical comparison of simulation results with real experiments.

Quantitative models can be deterministic or stochastic. In a deterministic model we assume that state values of the system could be treated as continuous and allow infinitesimal changes. This assumption is true when the concentrations of system components are relatively high and the frequency of reactions is significant. In this case the behaviour of the system is deterministic and predictable. When the number of molecules is low, reactions are infrequent and occur at random times. In this case a stochastic model can be constructed, which describes the probability distribution of state variables and the evolution of such distributions with time. When the concentration of the system elements increases to a reasonable value the results of deterministic and stochastic simulations should be identical, which gives the ability to investigate behavioural patterns with ODEs and then, if needed, use stochastic simulations to account for low concentration or noise within the system [3].

System of ordinary differential equations (ODEs) is one of the most popular deterministic quantitative modelling techniques at present. In this case each equation describes the evolution of a system variable (metabolite or protein concentration etc.) as a sum of conversion rates described with empirical laws (mass law, Michaelis law etc.) [5].

In this work we going to apply some of the techniques discussed above to obtain more rigorous understanding of the functioning and structure of one of the most intriguing subsystems of biological objects: their internal clocks.

1.2. Circadian Clocks

Nearly all eukaryotes and some prokaryotes have evolved a circadian clock to be able to adapt to daily changing environmental conditions [6]. The circadian clock generates internal ~24 hours biological rhythms to synchronise diverse physiological processes with the external day/night cycle, and this provides a selective advantage for growth and survival [7].

The circadian clock is capable of maintained sustainable oscillations under constant conditions. It is important, that clock is sensitive to external cues like light and temperature and, thus, able to adjust the internal rhythm to the variable conditions by entrainment. The clock is known to be temperature compensated, which means that the deviations in the temperature over some specific range do not significantly affect the main characteristics of the clock, such as period and amplitude [8]

Studies of the circadian clock in different organisms over the last thirty years have revealed the principle structure and regulatory properties of circadian clocks, which will be reviewed in the following sections

1.2.1. Molecular Mechanisms of the Circadian Clock

Classically, the clock is considered to consist of three main parts: input pathways, a core pacemaker and output pathways (Fig. 1.1) [9]

Input pathways maintain the synchronization of the core oscillator with the external daily cycle. This happens through specific regulation of some intracellular processes, like transcription or protein degradation, mainly by light. Living organisms perceive light by means of their photoreceptors which initiate the signalling cascade resulting in activation or inhibition of a particular clock component [10-13].

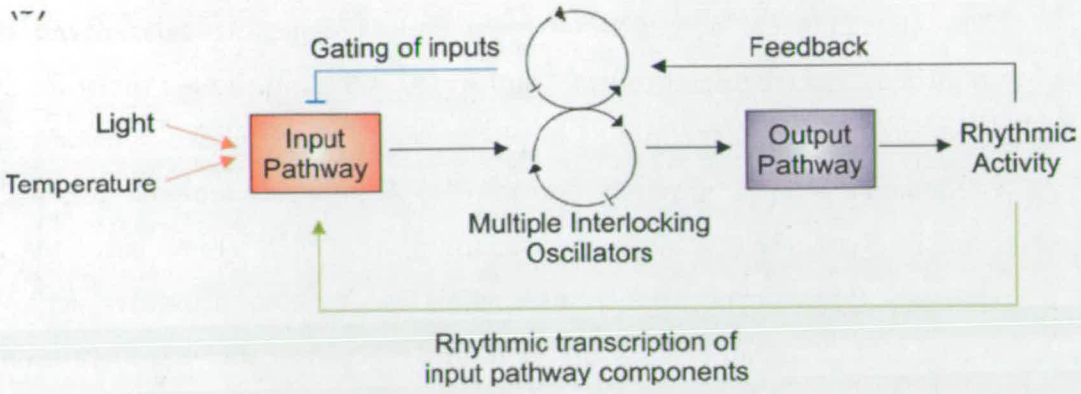


Figure 1.1 The current paradigm for the composition of molecular circadian clock (borrowed from [14])

The principle feature of the circadian clock is a core oscillator that generally comprises 5-13 genes that are rhythmically expressed [9]. These genes are organized in an interlocking structure of negative and positive feedback loops, with varying complexity peculiar to a given organism. It is significant that at the heart of this structure there is a simple transcriptional negative feedback loop, where the protein represses the expression of its own gene. It was shown by the mathematical modeling that the negative feedback loop with sufficient delay is capable of producing sustained oscillations [15, 16]. Nevertheless, the real pacemakers have far more complicated topologies and may include positive feedback loops that make the oscillatory behaviour more robust to external perturbations [17]. During the last years a growing amount of evidence suggests that for correct clock function the transcriptional circuit is complemented by a number of post translational events [18]. Interestingly, there is no or small homology between the transcriptional regulators of the fungal, plant and animal clocks, which means that the clock may have evolved independently several times. However, the principle topology of the network remains more or less similar [19], and post-transcriptional regulators are shared across taxa. Orthologous protein kinases and phosphatases in a similar way regulate the length of the circadian period in both vertebrates and *Drosophila* [20], while the highly conserved Casein Kinase II (CKII) is involved in circadian clock function in plants, fungi and animals [21].

The physiological adaptation to the daily changes is carried out through multiple clock-regulated output pathways, including those crucial for survival and increased fitness. Studies in plants revealed the phase–specific transcript abundance of approximately one third of *Arabidopsis thaliana* genes [22, 23], which gives the grounds for assuming that the clock regulates at the transcriptional level, and is responsible for defined promoter activity with respect to circadian cycle. This regulation is realized through highly conserved circadian clock regulatory promoter elements, which provide the synchronized expression pattern for the genes involved in the same pathway [23]. In plants such a regulation has been proposed for the key metabolic pathways related to the light-dark transition, including photosynthesis, starch and sucrose metabolism, nitrogen and sulphur fixation, hormone signalling, and stress-response [22, 23]. Clock-regulated pathways in turn affect core oscillator performance, feeding back in various ways, so that the clock appears strongly interconnected with other aspects of cell function [24].

In the next section I will illustrate the principle circadian topologic structure with several examples of known clocks. I will focus mainly at the intracellular level, though *in vivo* clock cells can be closely coupled, especially in neurones [25].

1.2.2. Topological Organisation of Natural Circadian Clocks

Historically, the first clocks studied at molecular level were *Drosophila* and *Neurospora* circadian clocks. The basic structure of the core oscillator in these organisms comprises the autoregulatory negative feedback loop, where the positive elements (generally the heterodimers of PAS-domain containing proteins) induce the transcription of the negative elements. The latter, when accumulated, retard the activity of the positive elements, decreasing the rate of their own transcription [9, 26] (Fig 1.2). The important part of the above mechanism is the posttranscriptional and posttranslational regulation, allowing the fine tuning of the oscillatory behaviour through the variation of the amount of component and the production of essential delays [18].

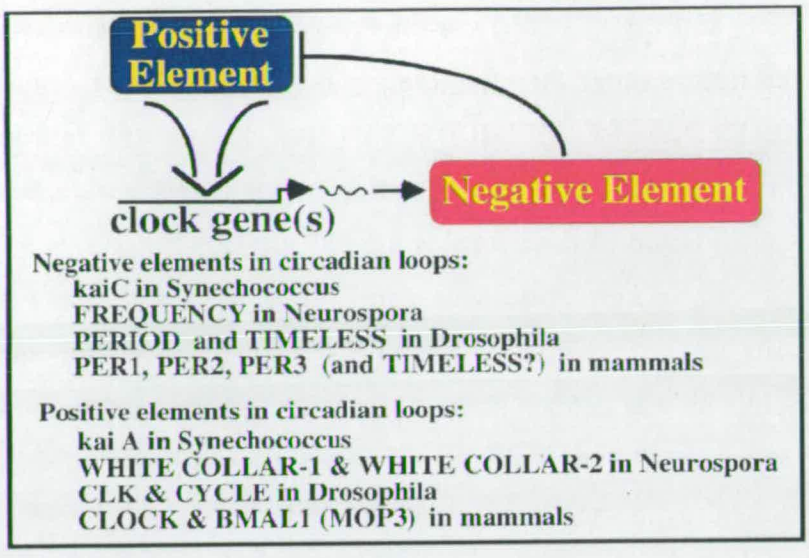


Figure 1.2 The principle structure element of the core pacemaker- negative feedback loop [9].

In *Neurospora* the positive elements, WC-1 (White Collar-1) and WC-2 (White Collar -2), form the White Collar Complex (WCC) which is able to activate the transcription of *frq* (*Frequency*) gene [6]. The FRQ protein in homodimer form inhibits the transcription of its own gene, 'freezing' the activating potential of the WCC complex. FRQ is also able to promote the accumulation of the WC-1 protein from available RNA, thus, producing the positive feedback which is considered to contribute to the increasing robustness of the core oscillator [27].

Both WC-1 and WC-2 are photoreceptors and, hence, transmit the light signals directly to the oscillator, rapidly increasing the rate of transcription from the *frq* promoter.

Progressive phosphorylation of the FRQ protein leads to its subsequent degradation, releasing the WCC from the repressed state. Meanwhile, for turnover of WC-1 the hypophosphorylated FRQ is needed. It makes a complex with WC-1 and promotes its export from the nucleus [28]. Thus, posttranslational regulation plays a significant role in closure of the central feedback loop [18]

The central feedback loop of the *Drosophila* oscillator generally resembles the structural organization described for *Neurospora*. The positive element CLK

(CLOCK) and CYC (CYCLE) are the PAS- domain containing transcription factors that in heterodimer form promote the transcription of negative elements TIM (TIMELESS) and PER (PERIOD) [9, 29, 30]. Nuclear PER represses the activation of *tim* and *per* promoters by binding to the transcription factors. The gradual phosphorylation of the cytoplasmic PER by DBT (DOUBLE-TIME) regulates the rate of its accumulation and contributes to the delays necessary for non-linearity [31]. The first mathematical model describing the molecular clock mechanisms was proposed for *Drosophila* and contained only a single gene, *per*. According to the model, the required time delay could be obtained from multiple phosphorylation of PER protein [32].

Earlier it was proposed that the *Drosophila* clock includes two more feedback loops both dealing with the expression level of the CLK transcription factor. Namely, Par- Domain Protein 1 (PDP1) feeds positively to CLK directly inducing its transcription. Another protein, VRILLE (VRI), feeds back negatively, repressing the transcription of CLK [33]. However, later it has been shown that neither PDP1 nor VRI control the expression of CLK and clock function as a whole. More likely they are required for the promoting the rhythmicity of flies locomotor activity and, thus, govern the clock output pathway [34]. The *Drosophila* core oscillator detects light by means of the photoreceptor CRY which is able to bind to TIM promoting its degradation; this mechanism underlies the resetting of the fly clock in response to light signals [35].

Orthologues for the fly clock elements were found in vertebrates, in particular, in mammals. The mouse clock comprises three PER proteins (PER1, PER2 and PER3); broadly they function in a similar way to fly PER, though phase of their expression may differ, like in the case of PER1 and PER2. The phosphorylation state of PER is regulated by casein kinase 1ε, which, as DBT does, provides the essential delays in PER accumulation. Similarly to flies, PER transcription is activated by the combined action of PAS-domain containing transcription factors CLK and BMAL1 (ortholog of *Drosophila cycle*). Cytoplasmic phosphorylation of PER by by several casein kinases regulates its stability. Heterodimer PER/CRY enters the nucleus and suppresses the BMAL1/CLOCK-

induced transcription of their own genes [36]. Another significant difference from flies is that CRY (a photoreceptor in *Drosophila*) takes a more active part in mammals, physically interacting with PER and promoting the downregulation of its expression. The second feedback loop is formed by REV-ERB and ROR proteins, whose expression is also under the control of CLOCK/BMAL1 complex. Both, in their turn, act on BMAL1, inhibiting and activating its transcription, respectively [36, 37].

Although the transcriptional gene regulatory network is assumed to be the universal clock structure, in fact oscillatory behaviour could be produced by any system that tends to leave the equilibrium state and return back to it. Any process whose product retards the process itself is capable of oscillation, and this has already been shown for a number of metabolic processes, for instance, glycolysis [38]. In light of this, it is not very surprising that in cyanobacterium *Synechococcus elongatus* the central clock pacemaker has been identified as working on the protein level. Three genes of *S. elongatus*, *kaiA*, *kaiB* and *kaiC*, comprise the circadian clock. KaiA protein supports the autokinase activity of KaiC protein, while KaiB attenuates this effect. The phosphorylation state of KaiC protein stably oscillates in a cell with 24h period even in the absence of constant transcription and translation of *kaiABC* operon. KaiC phosphorylation cycle showed persistent oscillations even when all three recombinant proteins were incubated *in vitro* [39]. Thus, post-translational modifications of KaiC including kinase/phosphatese and ATPase activities seem to account for general time- keeping of the period length. Since the biochemical clock has to be synchronized with transcriptional and translational events, the ratio of KaiC phosphorylation conjointly with a level of phoshorylated KaiC regulates the expression from kaiABC promoter. Hence, a negative feedback network topology could give rise to an oscillatory regime of the system despite the nature of the network elements and types of interactions between them [40].

1.2.3. The Arabidopsis Circadian Clock and Clock Modelling

Plants perceive light through multiple photoreceptors: red-light –sensing phytochromes (PHYs) and blue light–sensing cryptochromes (CRYs). The role of phytochromes using the example of PhyA and PhyB will be described in further detail below. The signalling pathway linking photoreception and the core pacemaker is not clearly understood, but the candidates for signal transmission like the PHYTOCHROME INTERACTION FACTOR (PIF) and PHYTOCHROME INTERACTION FACTOR-LIKE (PIL) DNA-binding proteins are under consideration. Although no evident orthologs of other clock genes have been identified among the *Arabidopsis* clock components, the structure of its core oscillator is likely to follow the current paradigm, being represented by interconnected transcriptional feedback loops [41].

The central loop consists of MYB-like transcription factors LATE ELONGATED HYPOCOTYL (LHY) and CIRCADIAN CLOCK ASSOCIATED1 (CCA1), which are supposed to share the same function, and TIMING OF CAB EXPRESSION (TOC1) (a member of the PSEUDO RESPONSE REGULATORs (PRR) family) [14, 42]. LHY and CCA1 repress the transcription by binding directly to the toc1 promoter. The TOC1 protein feeds back positively, activating the expression of LHY and CCA1. As for FRQ and PER, the accumulation of CCA1 is regulated by casein kinase CK2 [43]. Another feedback loop includes the PRR9 and PPR7 proteins, whose expression is induced by the CCA1 transcription factor. In their turn, they downregulate the expression of CCA1; hence, completing the feedback loop [44]. Several other proteins are thought to contribute to additional feedback loops, among them GI, ELF3, ELF4, LUX, etc. The interest of GIGANTEA (GI) increased after it was proposed to participate in a hypothetical gene function, predicted by a mathematical model [45]. EARLY FLOWERING 3 and 4 (ELF3 and ELF4) are both involved in photoperiod reception. Particularly, ELF3 mediates the gating of light response and light input to the clock [46, 47]. LUX ARRHYTHMO (LUX) is proposed to be regulated by CCA1/LHY in a similar manner to TOC1 and currently is among the key candidates to the core oscillator components [48] (Fig.1 3, A).

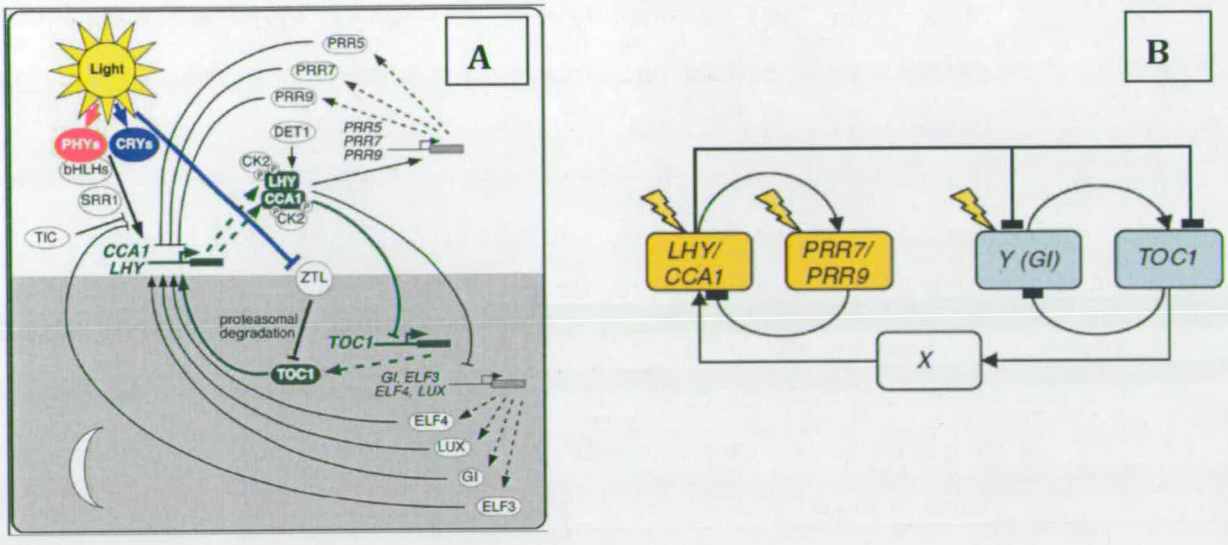


Figure 1.3 The Arabidopsis circadian clock.
A. The multiple components of the core Arabidopsis oscillator (borrowed from [41]).
B. The model of the Arabidopsis circadian clock (borrowed from [45]).

In the situation described above, with the great degree of complexity and ambiguous structural organization, it is very helpful to identify the minimal list of components with known regulatory relationships. After that, using the modelling approach, one could try to reproduce the behaviour of the complex system within a reduced one. Mathematical modelling allows understanding of the principle topologic structure needed to maintain the desirable function and deduce the properties of missing components that should be identified next and included in the model system.

The relatively small number of the clock genes with well-described interactions through positive and negative feedback loops, makes the circadian clock an appropriate object for ODE modelling. Several models of this type for *Drosophila*, *Neurospora* and mammals have been constructed during the last fifteen years [45 , 49-52]. The modelling process generally includes several steps. As a rule, the initial model for the system is quite simple and includes very few components. As I mentioned above, the first model for the *Drosophila* clock consisted of only one *per* gene [32]. After the model is constructed, its simulation reveals the similarities and discrepancies compared to the

observations that obtained from experiments. Each model usually aims to describe some certain phenomenon and often fails to match the set of other observations. The simple Goldbeter model was able to explain how the oscillatory behaviour rises from negative feedback regulation and time delay but was lacking a description of light entrainment. The later models, proposed for this system, were more detailed and included additional components. For instance, the model for interlocked feedback mechanism of *Drosophila* oscillator developed by Ueda at al could imitate the time courses for RNA and protein levels of the wild type and single and double mutants. It also reproduced the light- induced phase shift in the expression of clock components [53]. The interconnected combination of positive and negative feedback loops has been elegantly proved by mathematical modelling to account for the stabilization of the oscillations, with a particular attention to the possible mechanism of temperature compensation and gene dosage effects [54]. The first attempt to fit the experimental data with a simulation was realized by Forger and Peskin, who developed a detailed mathematical model of the mammal circadian clock [55]. Later, they included the stochastic component to account for the molecular noise, which made the model more precise [56].

In this way, in the *Arabidopsis* clock studies a major contribution in understanding the structure of the core oscillator has been accomplished by the mathematical modelling performed by James Locke[45, 52]. He first constructed a one-loop model with LHY/CCA1 and TOC1 connected as a negative feedback structure. Simulation and analysis of this simple model compared to emerging data revealed a necessity for a defined delay between the activation of TOC1 and its action on LHY/CCA1 transcription. To account for this, the non-identified component 'X' has been proposed. The candidates for this component are currently under examination. Although the single-loop model could sufficiently describe the behaviour of clock elements in the wild type, it failed to match the mutant data. This entailed the introduction of the second loop with a hypothetical element 'Y'. The GI gene has been suggested as a candidate for this component [52] (Fig 1.3, B). Finally, the model was revised and expanded so that it contained three loops, with the PRR7 and PRR9 proteins involved in the third one [45].

1.3. From Systems to Synthetic Biology

According to the above discussion, the identification of the main oscillator in the circadian clock system is not a straightforward process. First of all, oscillation is one of the most common types of behaviour of dynamic systems and could be demonstrated by networks with different topologies in many cellular functions. Thousands of genes are rhythmically regulated by the clock, but few of them contribute to clock function. To be able to prove most rigorously that a selected part of the cellular network is sufficient to form a pacemaker, one would have to recreate part of the network and demonstrate its ability to oscillate with a predictable period at least, if not also the other circadian properties. Though this has been achieved only for the cyanobacterial system described above, this points the way to the further progress in the area. Therefore, in systems biology where the emergent property of networks is the main object of research, synthetic biology becomes as important as *in vitro* investigations in biochemistry. From this point of view synthetic biology is a true hypothesis-driven science, where a hypothesis in the form of a dynamic or stochastic model of the synthetic network is formulated and then tested and adjusted by *in vivo* experiments.

Synthetic biology is a relatively new area of research, which combines science and engineering in order to design and build new biological functions and systems. It also refers to the re-design and fabrication of already existing biological systems for a better understanding of their principal structure and functioning.

As in the contemporaneous discipline of systems biology, synthetic biology also focuses on systems properties of the network rather than on its individual components, and uses the same respective tools of modelling, simulation and comparison to experiment as systems biology does. The main difference is in the generally engineering viewpoint of synthetic biologists. The emphasis on taking parts of natural biological systems, characterizing and simplifying them and eventually using them as building blocks of an artificial, engineered biological system is a characteristic feature of synthetic biology [57, 58].

Many useful insights from existing engineering fields are applicable to synthetic biology. Standardisation, abstraction, modularity, predictability, reliability and uniformity refer not only to electrical and mechanical engineering but also to the creation of artificial biological systems.

The basic principles of synthetic biology are modularity and hierarchy, so that every system can be considered as a set of modules, embedded in a complex system. By analogy with computer engineering, modules (complex pathways, integrated circuits) consist of devices (sets of biochemical reactions, protein-protein interactions), which in turn consist of small parts (DNA, RNA, proteins, metabolites, etc). The connection of these modules to each other provides unlimited ability to extend and modify the behaviour of the system in a predetermined fashion [59].

The synthetic design approach mainly implies the following steps (Fig.1.4), which are common in most of engineering projects. It starts with the development of a hypothetical 'design principle' for a desired phenotype accompanied with a prototype mathematical model. Both of these processes are inspired by physical and mathematical insight. Simultaneously, the components of the network with appropriate characteristics should be identified and implemented using biological elements. Finally, the network is reconstituted inside the cell to test the system's behaviour. Normally, it reveals numerous inconsistencies compared to the model that should be taken in consideration during further revision of the model [60]. However, it is hoped that better characterization (especially kinetics) of components will enhance the reliability of synthetic biological designs.

Building of artificial biological systems is usually accompanied and complicated by a unique set of problems, concerned with a major uncertainty in biological science. The lack of detailed kinetic parameters prevents the precise prediction of the network behaviour, on one hand, and possible interactions of the synthetic network with the rest of the cell, which are even less well characterised, on the other hand. This causes additional difficulty [60].

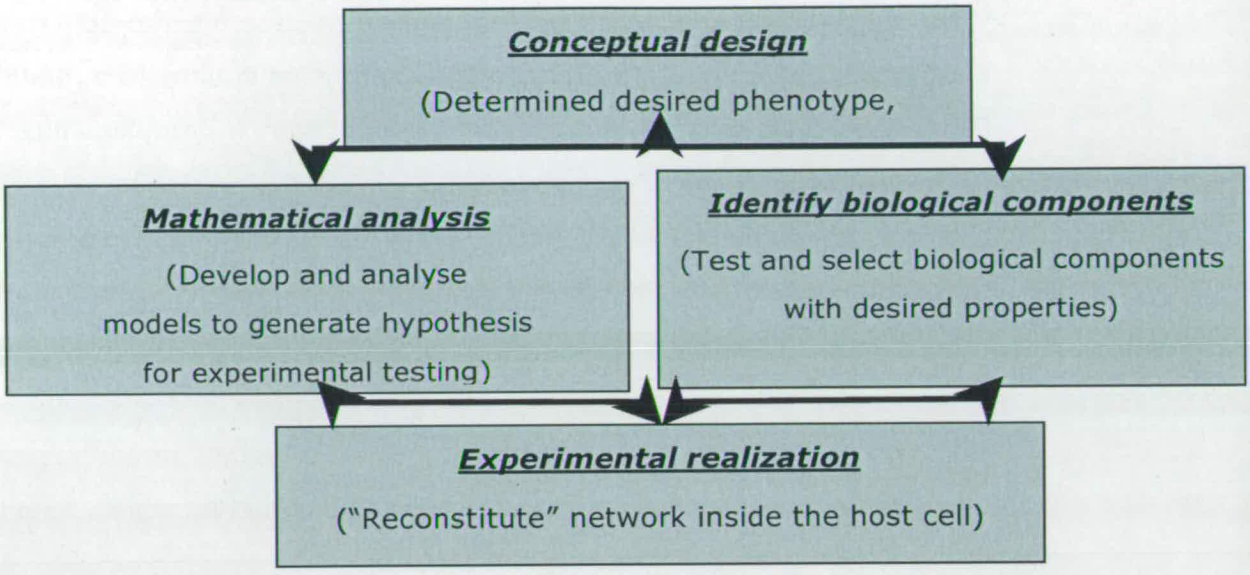


Figure 1.4 The main steps involved in the process of the creation of a synthetic network.

To overcome this uncertainty sometimes it is rational to use natural genetic logic modules, such as bacterial operons. These imply the set of genes which participate in a coordinated manner in some physiological response, being expressed from the same promoter [61]. As a rule, such subsystems have well-characterized biochemical and kinetic properties, which make them real biological devices with, increased predictability when used as components of the complex system. Another serious advantage of using bacterial operons is the minimizing of cross-talk with the host cell, especially when the host is a eukaryote. Good examples of such natural logic modules are the *E.coli* subsystems, namely the *lac* operon, which consists of three genes that participate in lactose metabolism, and the repressor-operator Tet-system responsible for tetracycline resistance. In synthetic designs they appear not in their intact form but in a form of modified gene regulatory modules with predictable response to the corresponding effective molecule (lactose or tetracycline). Many transcription regulatory networks have been engineered on the basis of these components so far, such as inducible switches, positive and negative feedback loops and oscillators [62-67].

With the increasing complexity of engineered systems the importance of mathematical modelling is coming to the forefront as it allows the prediction of the possible behaviour of the synthetic network before it is implemented *in vivo*. Phenomenological models comprise the detailed description of individual network components and the mechanisms of the biochemical reaction that those components have to undergo within the designed network. They predict the network's dynamics and suggest the structural modifications leading to the desired behaviour [61]. Bifurcation analysis applied to the ODE models serves as a very useful strategy for identification of the regions in parameter space where the system behaves as expected and the parameters whose values could be easily tuned to attain these regions [63, 68].

However, a phenomenological model will be unable to account for the heterogeneity at the population level that rises from the gene expression noise. The effects of the probabilistic nature of biochemical reaction appear more important when one has to deal with small molecular numbers. Thus, after the system is implemented *in vivo*, the revised model should comprise the stochastic, deterministic and phenomenological approaches to describe the network behaviour with a good level of accuracy [61].

As we have in mind the design of an artificial circadian clock, in the next subchapter I will present in more detail the synthetic oscillators that have been developed to date.

1.3.1. The Design of Intracellular Synthetic Oscillators

Understanding of the design principles of simple circuits is supposed to help in gaining insight into more complicated system dynamics. The first synthetic oscillator was a 'repressilator' constructed by Elowitz and Leibler [63]. It was constructed as a ring network of three repressors connected consecutively so that each of them repressed the transcription of the subsequent one within the network. Mathematical analysis revealed that not all possible parameter values led to oscillation; moreover, by altering the particular parameter the likelihood for desired behaviour could be distinctly increased. In this particular case, the performance was improved by minimizing promoter 'leakiness' and shortening

protein half-life [63]. This system was not designed to be synchronised and the oscillatory dynamics has been observed only on the single-cell level; nevertheless, the network evidently demonstrated sustained oscillations with a period of 150 minutes.

Another synthetic oscillator: the so-called 'Atkinson clock' could display periodic behaviour with a much longer period (600 minutes) although the oscillation was damped. The desired behaviour was demonstrated on the population level, with synchronization throughout the experiment [68]. The mathematical model for this system was proposed before the actual *in vivo* implementation. It predicted the possible dynamic behaviour of the system and as well as possible ways for its improvement [45]

An example of a synthetic oscillator with several main hallmarks of a natural clock, namely temperature compensation and close interaction with metabolism, is the construction, proposed by Fung *et al* [60, 69]. This 'metabolator' consists of a flux-carrying enzyme network with negatively and positively regulated expression of key enzymes and demonstrated a 45 minute period of oscillation [69].

Recently, a synthetic Tet based mammalian oscillator with a period of about 200 minutes was created. Its principle design distinction is a sense-antisense mRNA interaction which provides the negative feedback loop[67]. The oscillator also includes autoactivation of sense transcription, thereby increasing the robustness of the oscillations. The phenomenological mathematical model predicted that the oscillatory characteristics of the network would depend only on the relative concentrations of the components, and that was supported by the experiment. Further model revision was performed based on the statistical single-cell data and resulted in a semi-quantitative stochastic model [67].

1.4. Input pathway: Phytochrome Photoreceptors in Plants and Synthetic Photoreceptors

The synthetic oscillators presented above are not able to substitute real clock oscillations, for they all lack the main part of every clock system: the light input,

which allows synchronization of the internal oscillator with the day/night cycle. To recall, in living organisms light is perceived through special photoreceptors that mediate the light regulation of several thousand genes.

In plants, photoreceptors are represented by four known protein families: phytochromes, cryptochromes, phototropins and flavin-binding proteins of the FKF1 family [10]. Phytochromes in *Arabidopsis* are encoded by the multigene family PhyA-PhyE and represent chromoproteins comprising a polypeptide and a covalently linked tetrapyrrole chromaphore.

Each phytochrome can exist in two interconvertible isoforms: first synthesised in Pr form (biologically inactive, absorbs red light) it can be easily and reversibly converted into Pfr form (biologically active, absorbs far red light) [70]. The Pfr-form moves to the nucleus, which suggests a direct signalling pathway from phytochromes to target gene promoters [71] (Fig 1.5).

In dark-grown seedlings the most abundant phytochrome is PhyA, while in sunlight PhyB prevails due to more rapid PhyA degradation [72].

The properties of plant phytochromes and their ability to bind with other signalling proteins, e.g. with PIF3 (a member of the helix-loop-helix bHLH superfamily) have been investigated not only in plants [73, 74], but also in yeast [75]. Yeast results show evidence of noticeable distinctions in kinetic properties of phytochrome reactions in comparison with plants, e.g. dark reversion of phytochromes in plants and yeast has different rates [76].

For the purposes of synthetic circuit design, plant or cyanobacterial photoreceptors could be incorporated into organisms, which do not possess their own photoreceptors. One example of such a synthetic biology application is the project of The international Genetically Engineered Machine competition (iGEM) participants from the University of California in 2005, where a bacterial system was designed that could be switched between different states by red light. In this system the *Synechocystis* phytochrome Cph1 was fused to the *E. coli* histidine kinase domain which allowed a lawn of bacteria to function as a biological film [77].

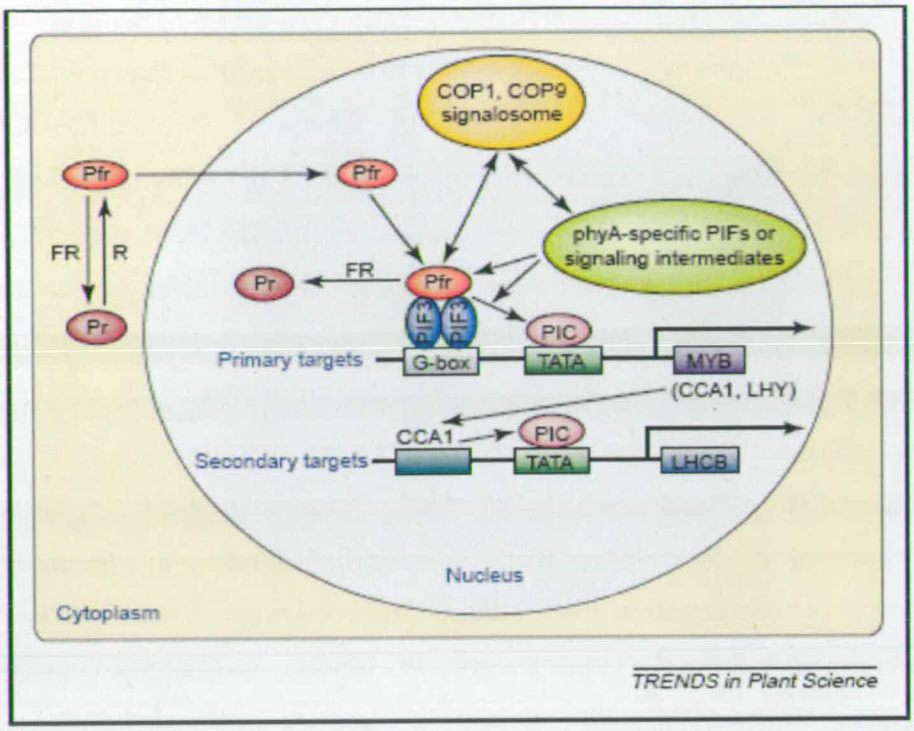


Figure 1.5 Mechanism of phytochrome A control of gene expression.
Diagram borrowed from [78]depicts the possible mechanism of phytochrome A control of gene expression. Upon the Red light induced photoconversion Pfr form of PhyA moves to nucleus where it interacts with PIF3 protein and activates the expression of primary target genes.

The first yeast transgene light-inducible system was created by Shimizu-Sato *et al* in 2003 [66]. It demonstrated many of the advantages of inducible systems, including low background expression, high inducibility, reversibility and dose-dependence. All these desirable features coexisted with non-toxicity and a lack of pleiotropic and unanticipated effects, which are inherent properties of chemically inducible systems. The system is based on the properties of the plant phytochrome B photoreceptor (PhyB), which reversibly changes its conformation in response to red (λ_{max} = 660 nm) or far-red light (λ_{max} = 730 nm). The far-red light absorbing conformer (PhyB Pfr) binds to the phytochrome interacting factor 3 (PIF3) proteins, whereas interaction between the red light absorbing conformer (PhyB Pr) and PIF3 is much less efficient. In the novel system, PhyB and PIF3 are expressed as chimeric proteins, fused to the DNA-binding (GBD) or the transcriptional activator (GAD) domain of the

GAL4 transcription factor, respectively, giving a typical two – hybrid interaction assay.. The *cis* component of the system is the *lacZ* reporter gene controlled by a GAL4-responsive artificial promoter. In darkness, PhyB-GBD binds the promoter, but does not induce transcription. Red light illumination converts PhyB into the Pfr form, therefore facilitating PhyB-PIF3 interaction, which recruits PIF3-GAD to the GAL4-dependent promoter resulting in the activation of transcription. Subsequent far-red light illumination coverts PhyB Pfr to Pr and this is followed by the dissociation of the PhyB-GBD – PIF3-GAD complex and abrogation of transcription. The authors demonstrated the dose-dependent response of the system and the dynamics of photoreversible activation of the *lacZ* reporter gene, derived from quantitative liquid culture assays [42]. Recently, another genetically encoded signalling system based on PhyB – PIF3 interaction, with different chimeric proteins binding, has been successfully used for photoswitching of actin assembly through the Cdc42-WASP-Arp2/3 pathway in *E.coli* [79]

1.5. Diurnal Regulation of Starch Metabolism as a Candidate Clock Output

Synthetic systems typically control simple reporter genes, which allow detection of the strength and the amplitude of the system's response to external signal. Natural systems have much more complicated outputs.

As introduced above, the adaptation of a living organism to the daily changing environment happens through multiple clock-regulated output pathways. For successful growth and development it is often crucial that certain physiological actions match the actual time of the day/night cycle. One of the most pronounced examples of such a connection is the diurnal rhythmicity of polysaccharide storage in plants. Plants accumulate approximately half of the CO_2 fixed during the light period as starch to buffer against diurnal carbon fluctuations. Remobilisation of the sucrose from starch during the night prevents plants from growth-inhibiting carbon depletion. Starch metabolism as

a whole thus appears as a highly significant aspect of the plant life and, accordingly, is a matter of complex regulation realized at several levels [80].

Starch is the main storage polysaccharide for plants, algae and cyanobacteria [81]. It has a complex semicrystalline granule structure composed of two distinct fractions: amylopectin and amylose. The former consists of medium-sized 1-4 glucans organized in a 3D helical complex by 1-6 linkages. The latter is typically referred to as long 1-4 linked linear glucans. The ratio of these fractions in the overall starch granule composition, and the ADP-glucose-based biosynthetic pathway Figure 1.6) localised to chloroplasts, comprise the main properties that are well conserved through all the Chloroplastidae [82].

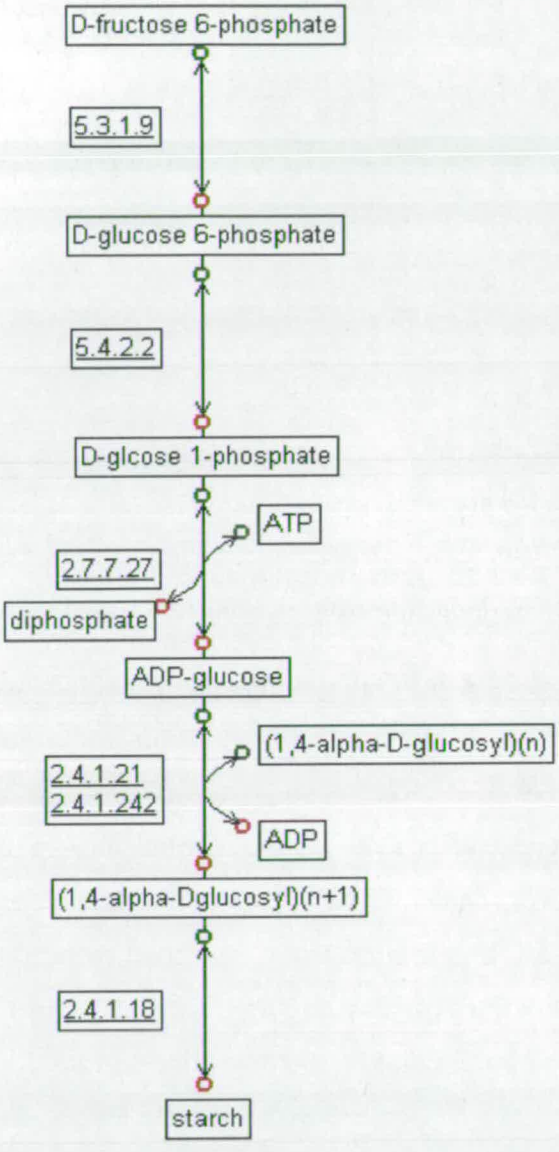


Figure 1.6 The pathway diagram of starch biosynthesis.

EC numbers correspond to the follwing starch biosynthetic enzymes: 5.3.1.9 Glucose-6-phosphate isomerase; 5.4.2.2 Phosphoglucomutase; 2.7.7.27 ADP-glucose phosphorylase (AGPase); 2.4.1.21 Starch synthase, 2.4.1.242 Granule-bound starch synthase; 2.4.1.18 Starch branching enzyme

On the transcriptional level the expression of genes for the enzymes of starch metabolism is regulated by the integrated inputs of the circadian clock, light/dark signals and sugars [83, 84]. Transcriptomes of both the synthetic and degrading *Arabidopsis* pathways show evident diurnal patterns [85, 86]. In particular, nine genes involved

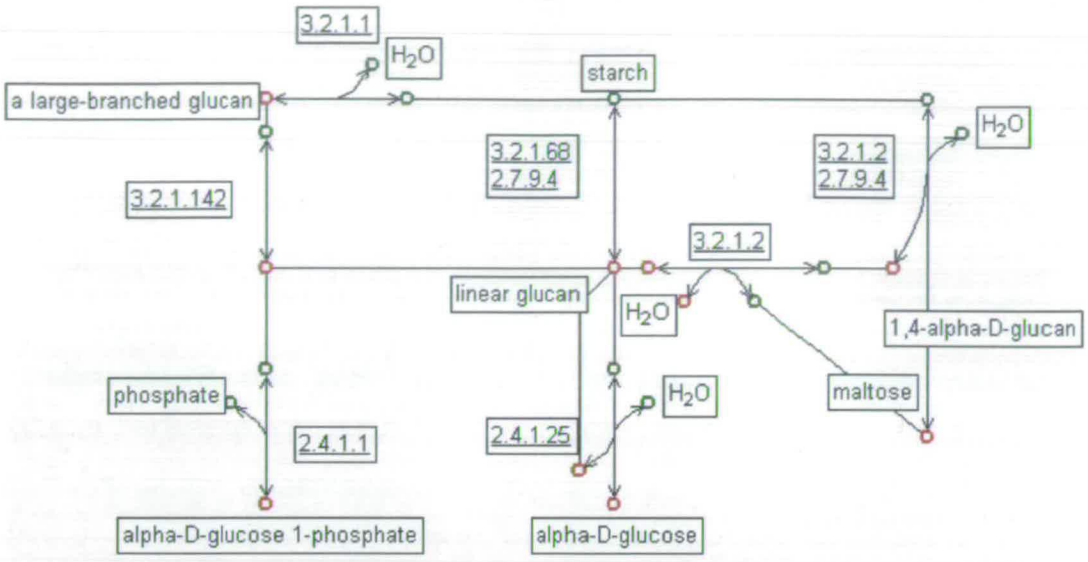


Figure 1.7 The pathway diagram for starch degradation.
EC numbers correspond to following starch degrading enzymes: 3.2.1.1 Alfa-amylase; 3.2.1.142 Pullalonase; 2.4.1.1 Starch phosphoylase; 3.2.1.68 Debranching enzyme; 2.4.1.25 Disproportionating enzyme; 2.7.9.4 Glucan, water dikinase; 3.2.1.2 Beta amylase

in starch catabolism (Figure 1.7) show a synchronized profile with the maximum expression level at the end of subjective light, and minimum at the end of subjective darkness [21, 22, 87]. Among them are the maltose transporter MEX1, starch dikinases (GWD, PWD) and phosphorylases SEX4 [83]. Their circadian patterns, however, could be significantly modified during the extended night treatment, when the inputs from negative carbon balance become dominating. [83]. The length of the day and light intensity jointly define the rate of starch biosynthesis, and accordingly, the degradation rate during the subsequent night [85, 88].

However, the protein abundance of starch enzymes in *Arabidopsis* and their activity do not evidently demonstrate causal links with the diurnal transcriptome profiles, staying relatively flat throughout the light-dark cycle [85]. For rapid adjustments they are likely to undergo a multi-step post-translational regulation. Among the several phenomena described are allosteric and substrate-product regulation, reversible protein phosphorylation and redox potential regulation [89]. The latter seems the most significant for the events

spatially related to the chloroplast, as it connects light, sugar and pH signalling in one highly organized scenario [83, 84]. Photosynthetic electron transport chain is believed to be involved in the regulation of the chloroplast- specific pathways. Among them are the Calvin cycle, amino acids and lipids synthesis.

The regulation is realized through the transfer of electrons from the Photosystem 1 via ferredoxins/thioredoxins to the target proteins which are becoming activated by the reduction of their disulfide bonds [90]. Sugars, namely sucrose and glucose are also involved in redox signalling. For instance, glucose feeding leads to increase in $NADPH/NADP^+$ ratio, which affects plastidial redox regulation [90]. It is not surprising, therefore, that redox regulation has already been reported to be a determinant for the main rate-controlling steps for both starch biosynthesis and degradation: ADP-glucose phosphorylase, glucan water dikinase 1 (GWD1, R1 enzyme) and β–amylase (TR-BAMY) [91-93]. Redox –dependent modulation of AGPase is responsible for channelling of incoming sucrose towards the starch, thus, the process could be simulated specifically by sugar sensing. On the other hand, the degradation of semicrystalline starch granule may require the highly coordinated association of several enzymes on the surface. This is triggered by decreased reduction state of thioredoxins, which in turn links starch degradation to the light and carbon status of the plant leaf [90].

However, the overall complexity of the regulation of the starch metabolism, briefly introduced above, appears to be largely specific to higher plants with their tissue and organ compartmentation and, thus, with many concurrent events in distinct locations. The photosynthetic unicellular red alga *Galdieria sulphuraria* has been studied in respect to the redox regulation properties of several enzymes involved in the Calvin cycle [94]. Intriguingly, despite high sequence similarity of the *G. sulphuraria* enzymes to those of other photosynthetic organisms, little or no redox regulation has been found. It appears that level of the redox regulation itself could be a matter of adaptation and for the land plants with their highly organized tissue and organ specificity,

rapid adjustments to physiological and environmental changes could be of more importance than for marine unicellular algae.

From this point of view it seems attractive to reduce and simplify the system picking out the essential mechanism of regulation on the basis of a unicellular photosynthetic cell.

1.6. *Ostreococcus tauri* – a Model System for Molecular Plant Science

Although synthetic biological engineering is growing to be an accepted tool for unveiling the principle features of relatively small gene regulatory circuits, it is virtually useless for the problems regarding huge multicomponent networks, like metabolic ones. If one conceives to track the clock–metabolism regulatory relationship using the modelling approach, he rather needs to deal with the natural reduced and simplified object retaining all the general properties of the more complicated ones. Here we come to the so-called 'model' organisms.

Unicellular green algae share the main metabolic functions with higher plants. Moreover, the less complicated cellular infrastructure and relatively simple experimental protocols make them advantageous for scientific inquiry, including, for instance, the circadian clock related investigations, where the liquid culture could be easily synchronized with the 12/12 light-dark photoperiod. *Chlamydomonas reinhardtii* has been used as a model organism for plant biology for the last half the century [95]. Recently, four Prasinophyceae algae, *Ostreococcus* spp *tauri* and *lucimarinus* and two *Micromonas species* have been sequenced [80, 96]. The latter are even more promising as models for basic plant investigations.

The smallest, free-living phytoplanktonic eukaryote, *Ostreococcus tauri* [97, 98] has an extremely simple cellular organization and very compact genome (12.5Mb) [96]. Regarding the clock, only two orthologs of Arabidopsis clock genes, namely CCA1 and TOC1, were identified as central clock genes (Fig 1.8). Thus, a minimal natural model, consisting of these genes interacting within a transcriptional feedback loop, might be considered for the clock function in *O.*

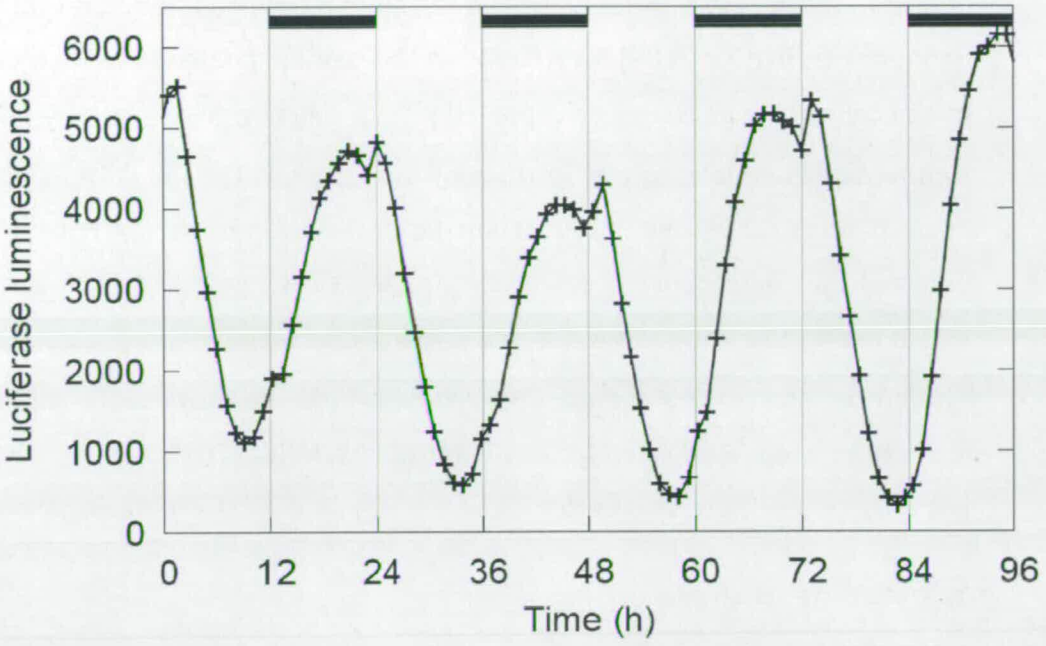


Figure 1.8 CCA1::LUC in light/dark cycles (LD 12:12).
Timeseries data are shown with luciferase reporter from François-Yves Bouget et al. (in press). The LHY7 line is shown in the 10/01/2008 experiment.

tauri. The dynamic modelling of the *O. tauri* circadian clock is currently being carried out in our lab by Carl Troein (personal communication).

Despite the overall genome simplicity, with respect to polysaccharide metabolism *Ostreococcus* demonstrates a complexity comparable with higher plants [99]. The ADP glucose–based starch biosynthetic pathway contains all the steps, characteristic of plants and Chloropycean algae such as *C. reinhardtii*, with multiple enzyme forms for each pathway step (Figure 1.5). Interestingly, *C. reinhardtii* seems to lack the redox regulation at least at the level of AGPase, as was shown in [100].

Ostreococcus tauri accumulates starch within only one granule that takes up most of the space within a single chloroplast. The partitioning and division of the granule coincide with the division of the chloroplast and the whole cell itself, which is subject to circadian clock control [101]. Therefore, cell division at the end of the light phase seems to dictate the local time of elongation and division of starch granule and, therefore, the overall timing for the starch synthesis and degradation cycle [99]. In *C. reinhardtii* the cell cycle is also clearly gated by the

circadian clock [102], influencing the starch content oscillations rather than the light dark transition [100]. Regulation of the storage polysaccharide metabolism by the cell cycle has also been reported for *Saccharomyces cerevisiae* [103].

Neither for *Ostreococcus* nor for *Chlamydomonas* has the full starch deprivation been demonstrated, even after prolonged incubation in the darkness [99]. It is believed that Ostreococcus produces the new starch granules from the old one, which makes the polysaccharide priming unnecessary [99].

Therefore, starch metabolism regulation could be noticeably different for unicellular algae compared to the higher plants, at least to the mature leaf tissue with stopped cell division cycles.

1.7. Outline of thesis

Understanding of the circadian clock mechanism and its downstream effects would be beneficial from various prospects including human health and agriculture.

This work, therefore, aims to get some insight into the principle mechanism of the clock functioning, specifically on the light input, core oscillator and output metabolic pathway, through the mathematical modelling techniques, which contribute to the powerful tool of Systems Biology.

The two first parts of the thesis focus on the development of the mathematical models for the synthetic biological systems: light input to the clock and core oscillator.

The kinetic deterministic model for light inducible Phytochrome-based system has been developed and fully validated by construction in yeast host, performed by collaborators from Ferenc Nagy's group, Szeged, Hungary (Chapter 2)

We used the same modelling approach to investigate two possible designs for the synthetic core oscillator, one is composed of yeast-derived components, and another contains the components of prokaryotic origin (Chapter 3). We found the second system more promising from the point of view of its ability to produce stable oscillations. The main analysis methods used there are the Forward and Inverse Bifurcation Analysis. They allowed us to find the

biologically plausible ranges for the parameters of our simplified system, that condition to the desired type of behaviour. We also examined the possibility of the entrainment of core oscillator by external forcing cycles provided by the modelled light input.

We applied a different modelling approach (quasi- steady state) to unveil the possible clock downstream regulatory mechanisms by the example of the starch metabolic pathway in model organism *Ostreococcus tauri* (Chapter 4). The model is based on literature – based knowledge about the pathway structure and demonstrates the predictive power with respect to the diurnal dynamics of starch content.

The following is the list of manuscripts published and in preparation:

1) A switchable light-input, light-output system modelled and constructed in yeast, **Oxana Sorokina**. Anita Kapus, Kata Terecskei, Laura E Dixon, Laszlo Kozma-Bognar, Ferenc Nagy, Andrew J Millar
   Journal of Biological Engineering 2009, 3:15 (17 September 2009)
2) Microarray data can predict metabolite levels in *Ostreococcus tauri*, **Oxana Sorokina**, Florence Correlow, David Dauvillei, François-Yves Bouget, Anatoly Sorokin, Igor Goryanin, Andrew J. Millar
   PLoS Computational Biology (submitted)

This thesis is also cited in:

"Inverse Eigenvalue Problems for Exploring the Dynamics of Systems Biology Models", J.Lu, *Advances in Applied Mathematics and Mechanics* (to appear).

Chapter 2. Modelling of circadian clock input

Our aim was to create and mathematically model an inducible gene expression system, which by analogy to [42] will be based on plant photoreceptor properties, but containing novel components that provide more stringent regulation and *in vivo* real-time detection of transcription in yeast colonies on solid media.

2.2. Introduction

All phytochromes (PhyA-E) in the model plant *Arabidopsis thaliana* are capable of light-dependent conformational changes, but interacting proteins have only been investigated for the two most abundant phytochromes (PhyA and PhyB)[75, 104, 105]. FAR-RED ELONGATED HYPOCOTYL 1 (FHY1) and FHY1 LIKE (FHL) proteins control the nuclear import of PhyA via specific interactions with the Pfr conformer [106, 107]. It follows that, besides the PhyB-PIF3 pair, other phytochrome-interacting protein combinations could be employed as the "light sensing" module of the expression system.

The light switch described by [6], translates light-dependent protein interactions into transcriptional regulation of a selected gene. Beta-galactosidase is the most widely used reporter gene in yeast; however, the protein has a half-life of more than 20 hours in this system [108] and it can be detected *in vitro* only. By comparison, the firefly luciferase has a 1.5 hour half-life in yeast [109] and luciferase activity (luminescence) can be monitored in real-time and *in vivo*, which makes this reporter a better tool for monitoring dynamic changes in transcription, as has been elegantly demonstrated recently through the monitoring of the cell-cycle and respiratory oscillations monitoring in agitated liquid yeast culture [110].

We used the luciferase quantitative readout to construct a mathematical model that matches the system's behaviour and predicts the molecular targets for future manipulation.

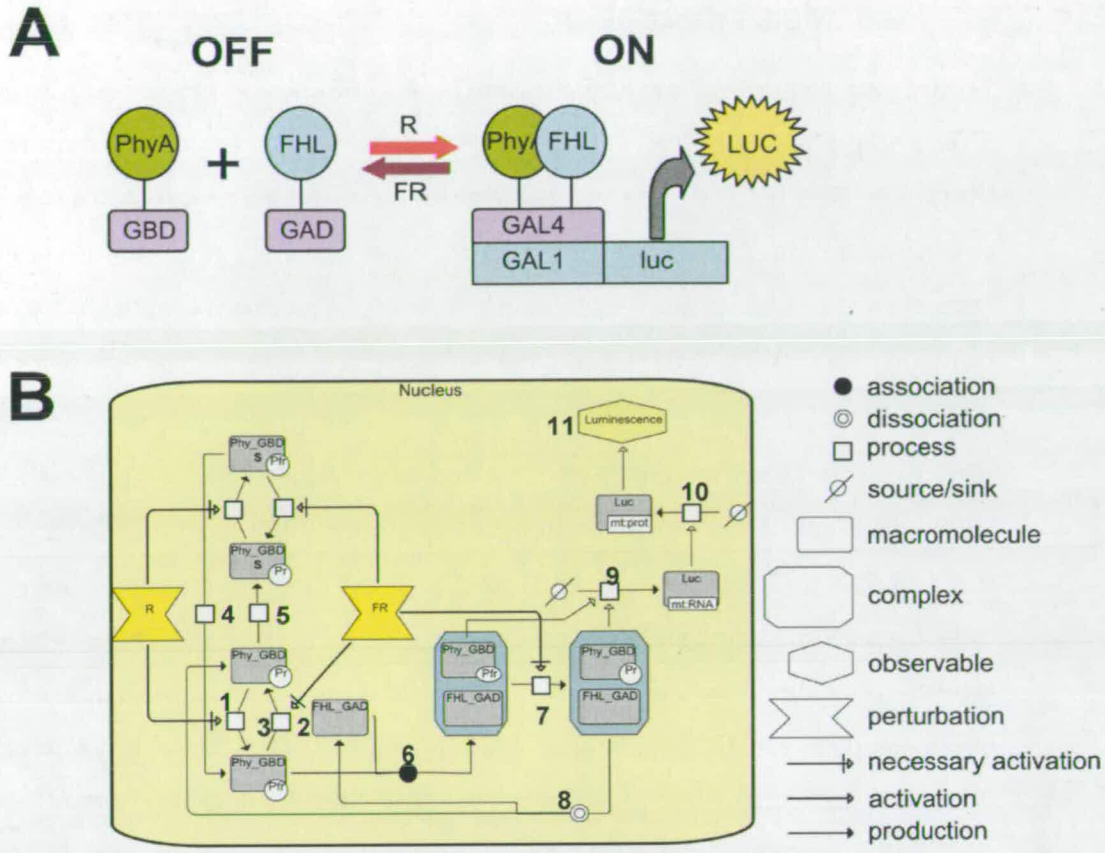


Figure 2.1 Light-responsive gene promoter system.

Target gene expression is activated by red light, when photoactivated phytochrome A (Pfr_GBD) interacts with PFL_GAD and recruits it to the target promoter with Gal UAS sites for transcription activation. Expression of the gene could be switched off by far red application, when photoinactivated phytochrome (PR_GBD) dissociates from FHL_GAD. (A).Schematic representation (B) SBGN representation[111] (See also Appendix A.1.for details). Presented reactions are:

***1* Phytochrome Pr-Pfr photoconversion by red light (R)**

***2* Phytochrome Pfr-Pr photoconversion by far red light (FR)**

***3* Pfr-Pr dark reversion**

***4* Pfr dissociation from SAPs (sequestered areas of phytochrome)**

***5* Pr association to SAPs**

***6* Phytochrome (Pfr_GBD) association with FHL_GAD**

***7* Pfr_FHL-Pr_FHL transition by FR**

***8* Intermediate Pr_FHL complex dissociation**

***9* Transcription of luciferase gene**

***10* Translation of luciferase mRNA**

***11* Luciferin- luciferase enzymatic reaction**

This chapter is based on work published in "A switchable light-input, light-output system modelled and constructed in yeast." [112].

2.3. Model description.

To use our regulatory system for synthetic biology we developed an ordinary differential equation (ODE) model of its function based on kinetic data from the literature and experimentally determined parameter values. The model describes all the known phytochrome properties (e.g. photoconversion, dark reversion, sequestration, etc), using yeast phytochrome data to provide a realistic description of the light-switch function.

Our principal model system (Fig 2.1, A, B) includes two chimeric proteins: phytochrome fused to the GAL4 DNA-binding domain (Phy_GBD), in the active (Pfr) and inactive (Pr) forms, and binding protein PIF3 (or FHL/FHY1) fused to the GAL4 activation domain (FHL_GAD). According to existing experimental data, the recombinant phytochromes are quite stable in yeast. Although the light lability of plant PhyA Pfr is well-described, no detectable difference was observed between the stability of the Pfr and Pr forms of oat PhyA over an 80 hour time period in yeast [113]; moreover, no significant decay in the total PhyA and PhyB amounts over 120 hours was reported by [76] and [114]. This provides the basis for assuming that our model proteins are present constitutively, so neither production nor degradation occurs in the model system. Two pools, Pool_Phy and Pool_PIF3, fulfil the mass conservation laws for Phy and PIF3.

In plants, the Pr forms of phytochromes are localized in the cytoplasm in the dark and are translocated to the nucleus in their Pfr form after light absorption [115]. In the yeast system, however, all fusion proteins are constitutively nuclear-localized due to the natural Nuclear Localisation Sequence (NLS) present in the GBD tag or the presence of the SV40 NLS motif fused to the GAD fusion partner. Therefore, in this system the only light-dependent event is the interaction of phytochromes with their corresponding protein partners. Taken together, these details give us reason to locate the interacting proteins and the processes of association and dissociation in the nucleus.

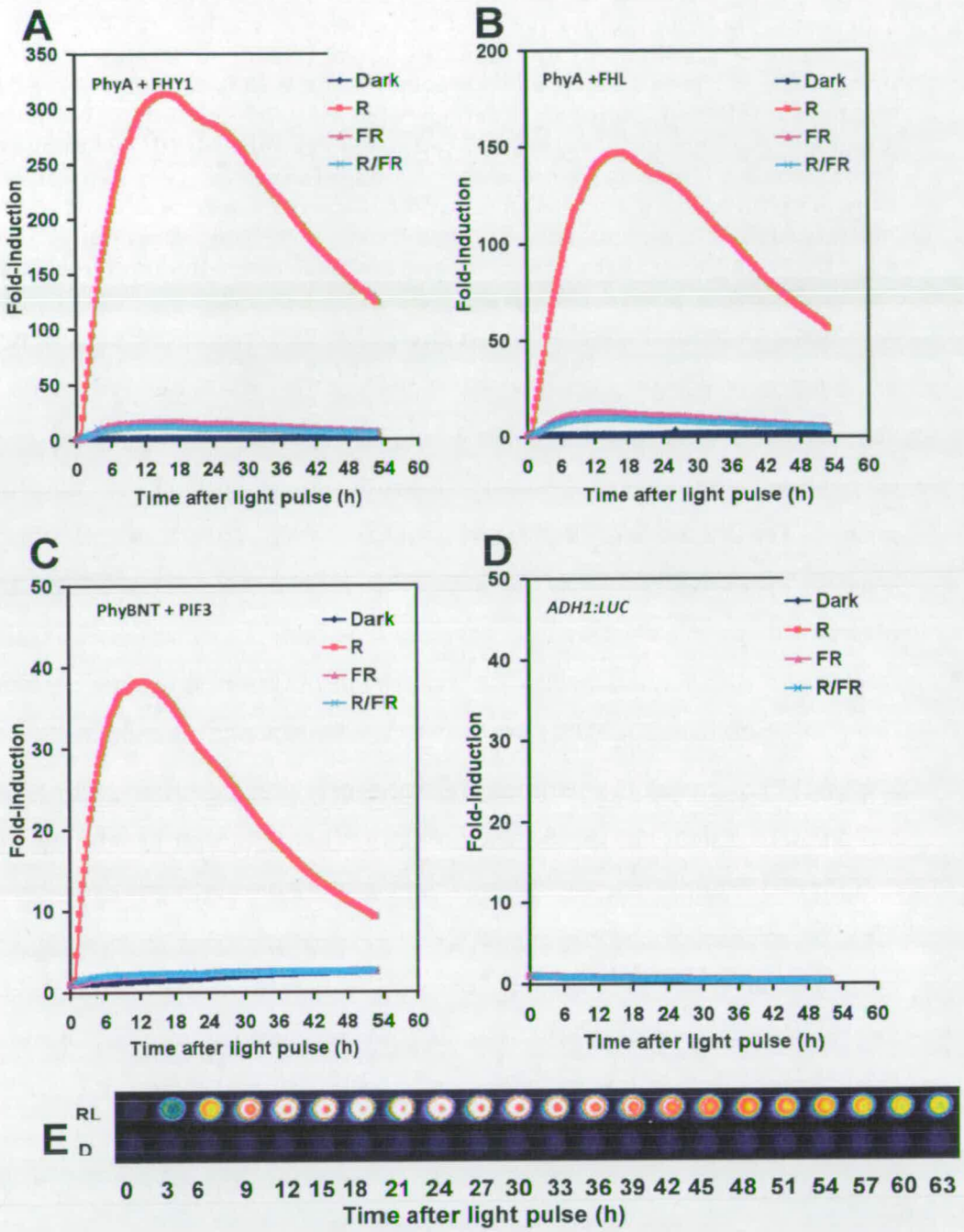


Figure 2.2 Light switchable light output.

Yeast cells harboring the GAL1:LUC reporter and expressing PHYA-GBD/GAD-kFHY1 (A), PHYA-GBD/GAD-FHL (B), or PHYBNT-GBD/GAD-PIF3 (C) fusion protein-pairs were grown in darkness to form patches (merged colonies) for two days at 30°C, treated with 2.5 mM luciferin and transferred to 22°C for 17.5 h. Separate yeast patches were irradiated with single red (R), or far-red (FR) light pulses, or with red pulses immediately followed by far-red pulses (R/FR), or were kept in darkness (Dark). (D) shows the control experiment when the same

protocol of light treatment has been applied to the yeast cells expressing the Luc (luciferase) marker under the constitutive promoter ADH1. (E) The real-time imaging of the yeast colony under the Red-light treatment. Luminescence values normalised to the pre-pulse levels are shown; time 0h is the start of the light treatment. Selected luminescent images of yeast patches used to obtain data in panel A. I: red light induced, NI: non-induced dark control, T0: last images before the light pulse. Consecutive luminescent images taken in every two hours are shown. Pictures are displayed in pseudo-colors: red-white or blue-black colors indicate high or low expression levels, respectively. The experiments have been done by Anita Kapus, Kata Terecskei and Laszlo Kozma –Bognar from Biological Research Centre, Szeged, Hungary.

Non-instantaneous kinetics of induction (Fig 2.2) prompted us to suggest the existence of two phytochrome pools: slow and fast. It has been reported that the sequestration of recombinant PhyA into the cytosolic SAPs (sequestered areas of phytochrome) in yeast has no dependence on light [114]. We, therefore, propose the presence of sequestered and free Phy pools (less and more easy to access, respectively) in the nucleus with a reversible interchange occurring between them. We assume that only the free pool is available for binding to its interaction partner, and, thus, the transition between slow (sequestered) and fast (free) pool is responsible for the shape of initial light response.

It is well known that the phytochrome photoconversion cross-section (σ) for Pr and Pfr forms depends on the wavelength of light. Red (approximately 660nm) and far red (approximately 730nm) light are the most effective for Pr→Pfr and Pfr→Pr photoconversions, respectively. Nevertheless, it is evident from the cross-section data [116] that the absorption spectra of the Pr and Pfr forms of Phy significantly overlap. This means that monochromatic light of biologically relevant wavelength (i.e. red) does not convert all the Phy to the Pr or Pfr form, but rather determines the specific distribution ratio of the forms in the total Phy pool. We thus have to account for the activation and inactivation of phytochrome by both red and far-red light, so that:

Photo_act = (Ka_R*RL + Ka_FR*FRL)*Pr_GBD

Photo_inact = (Ki_R*RL + Ki_FR*FRL)*Pfr_GBD

Exact values for Ki and Ka for the different wavelengths were adopted from [116]. In the model Pr ↔ Pfr transitions are applied for both associated and free form of the phytochromes.

For estimation of photoconversion rates we used the data for the photoconversion cross-section of Pr , Pfr and P and Pfr/P ratios at photoequilibrium of type -I phytochrome [116].

- Ka for R and FR were estimated according to the wavelength for 660nm R and 730 nm FR: Ka_R=4963e-6 ($m^2/\mu mol$); Ka_FR=35.53e-6($m^2/\mu mol$); Ki_R=743.9e-6($m^2/\mu mol$); Ki_FR=1701e-6($m^2/\mu mol$)
- Light intensity for the R from experiment conditions is 70 ($\mu mol * m^{-2} s^{-1}$)=>70*3600=252000($\mu mol * m^{-2} h^{-1}$)
- For FR, 80 ($\mu mol * m^{-2} s^{-1}$)=>80*3600=288000 ($\mu mol * m^{-2} h^{-1}$)

Dark reversion has been reported for PhyA and PhyB in yeast cultures [76, 114, 117]. According to these data, only a fraction of the total Pfr pool is subject to dark reversion (20-40 % of the total amount) with a half-life of 20-40 min. For simplicity in the current model we assume a single Pfr pool that is all dark reversible and has a longer half-life than the range suggested by Hennig et al [117]; however, the model is still in good agreement with the overall kinetics described in the literature [76, 114] .

We assume that dark reversion of the complex Pfr_PIF3 (Pfr_FHY1, Pfr_FHL occurs with the same rate. Therefore, both the photoconversion and dark reversion processes contribute to dissociation of the transcriptional activation complex.

Finally, for the PhyA_FHY1/FHL complexes, we assume the existence of an additional state, Pr_FHY1/FHL, that has the ability to activate transcription to some extent, as it has been previously demonstrated by [106]. Although the reference above corresponds to PhyA, in our experimental conditions PhyB demonstrated the same kinetics (Fig 2.2, C), so we assume the intermediate state for PhyB-PIF3 as well. According to our hypothesis, this complex is produced as an intermediate product of photoconversion of the Pfr_FHY1/FHL complex after FR exposure. Thus, we propose that Pr proteins that have previously been Pfr can interact with FHY1/FHL and activate transcription.

In summary, the model assumptions are:

1) Overall concentrations of Phys and PIF3/FHY1/FHL are constant,

2) Before the light impulse all the phytochromes are in the inactive (Pr) form, and sequestered in a slow acting pool. This might be related to their inclusion in SAPs (Sequestrated Areas of Phytochromes)-like structures similar to those observed by microscopy in cytosol [15]

3) The dark reversion rate is the same for the free Phy and the Phy-FHY1/FHL complex.

4) Photoconversion or dark reversion of the Pfr_FHY1/FHL complex creates an intermediate step (Pr_FHY1/FHL) with distinct kinetics, prior to dissociation of the complex.

5) The luciferase - luciferin subsystem is approximated as a steady state before light treatments.

6) The initial sharp decrease in luminescence following the application of luciferin is due to the diffusion of luciferin from the point of application on the yeast patch into the agar medium.

This enables the model to provide a good fit to conceptually similar systems with different interaction partners. For example, an adapted model fits Shimizu-Sato's data with good accuracy using parameter values derived from the literature [22, 29, 66, 118]. This model is simpler, for the main part because the slow LacZ degradation obscures the long-term kinetics (see the model equations and figures in Appendix A.3).

Mass Action kinetics were used to describe complex formation and dissociation, translocation, translation, and degradation. Transcription was described with a Hill function and the reporter enzymatic reaction follows Michaelis-Menten kinetics (see Fig 2.1, B and Appendix A.1 for the reactions presented).

The model equations for the PhyA_FHY1/FHL system are presented below:

$$\frac{d}{dt}Pfr\_GBD\_c = -Pfr\_GBDtrs - Darkrev\_c + Phyact\_c - Phyinact\_c \quad \textbf{(2.1)}$$

$$\frac{d}{dt}Pfr\_GBD\_f = Pfr\_GBDtrs - Darkrev\_f - Pfr\_FHLcplx + Phyact\_f - Phyinact\_f \quad \textbf{(2.2)}$$

$$\frac{d}{dt}Pfr\_FHL = +Pfr\_FHLcplx - Darkrev\_cplx - Pfr\_FHL\_fcon \quad \textbf{(2.3)}$$

$$\frac{d}{dt}\mathrm{Pr}\_GBD\_f = +Darkrev\_f + \mathrm{Pr}\_FHLdiss - Phyact\_f + PhyInact\_f \quad \textbf{(2.4)}$$

$$\frac{d}{dt}\mathrm{Pr}\_FHL = +Darkrev\_cplx - \mathrm{Pr}\_FHLdiss + Pfr\_FHL\_fcon \quad \textbf{(2.5)}$$

$$\frac{d}{dt}luc = luctrnbasic - luc\,\mathrm{deg} + luctrnind \quad \textbf{(2.6)}$$

$$\frac{d}{dt}LUC = +LUCtrl - LUC\,\mathrm{deg} - LUCS\,\mathrm{deg} \quad \textbf{(2.7)}$$

The equations (2.1)-(2.5) describe changes in concentrations of *all* the phytochrome components, where the terms *Pfr_GBD_c* and *Pfr_GBD_f* corresponds to active phytochrome in sequestrated and free form, respectively, and *Pr_GBD_c* and *Pr_GBD_f* correspond to inactive phytochrome in sequestrated and free form. *Pfr_GBDtrs, Darkrev_c, Darkrev_f, Phyact_c, Phyinact_c, Phyact_f, Phyinact_f, Pfr_FHLcplx* and *Pfr_FHL_fcon* correspond to reactions of dissociation of active phytochrome from SAPs, dark reversion of sequestrated phytochrome, dark reversion of free phytochrome, photoactivation of sequestrated phytochrome, photoinactivation of sequestrated phytochrome, photoactivation of free phytochrome, photoinactivation of free phytochrome, association of Pfr_FHL complex and dissociation of Pfr_FHL complex due to Pfr/Pr photoconversion. Also, *Pr_FHL_diss, darkcrev_cplx* correspond to dissociation of Pr_FHL intermediate state and dark reversion of the Pfr in the Pfr_FHL complex. Equations (2.6) and (2.7) correspond to the changes in concentrations of luciferase mRNA and

protein, respectively. Here, *luctrnbasic, luctrnind* and *lucdeg* correspond to the basal transcription of luciferase RNA, induced transcription of luciferase RNA and degradation of luciferase RNA. *LUCtrl, LUCdeg* and *LUC1deg* correspond to the translation of luciferase protein, degradation of the luciferase protein and luciferin dependent degradation of the luciferase protein.

Luminescence level is calculated according to the Michaelis-Menten equation:

$$LuM = k\_cat * LUC * \frac{Scyt}{Km + Scyt} * RLU \quad , \tag{2.8}$$

where k_cat is luciferase activity constant, Scyt is subsatrate (luciferin) concentration and Km is the Mychaelis constant for the luciferase- luciferin reaction.

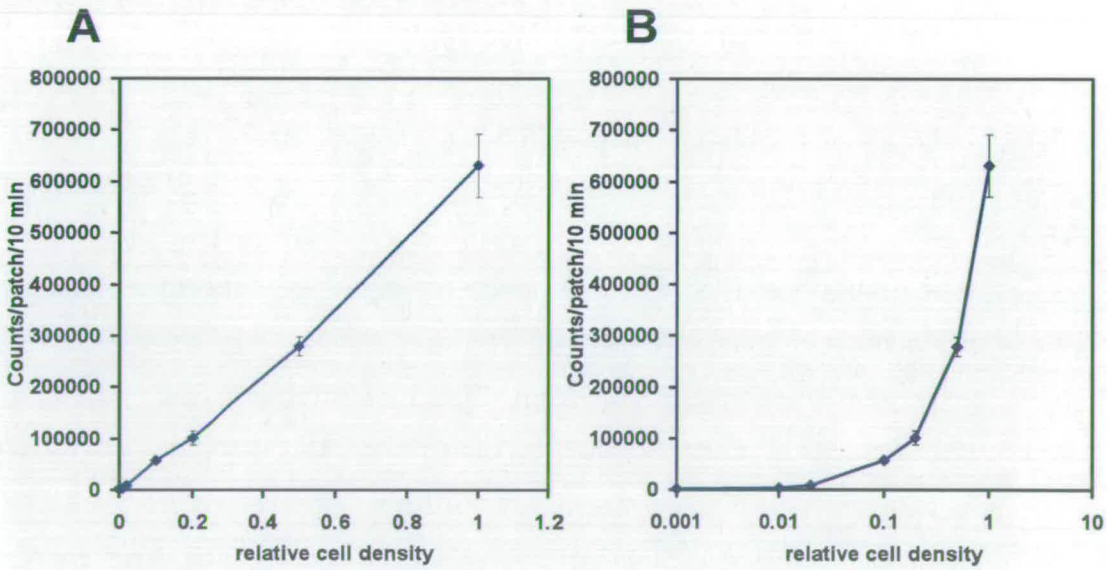


Figure 2.3 Effect of cell density on LUC luminescence intensity [112].
Yeast cells harboring the *ADH1:LUC* reporter were grown in liquid culture at 30°C for 16 h. Dilution series were prepared from the overnight culture and 20 µl of each dilution were pipetted on solid media and allowed to dry for 15 min at 22°C. The patches were treated with 2.5 mM luciferin and imaged subsequently. Average luminescence values of three replicates for each dilution were plotted against the relative cell density. Graphs with linear (A) or semi-logarithmic (B) scale are shown. Error bars represent standard deviation. The culture with 0.01 relative cell density had OD260=0.232 and the corresponding patches contained $8.2 \cdot 10^4$ cells. Yeast cells were counted by a Burker counting chamber. The experiments have been done by Anita Kapus, Kata Terecskei and Laszlo Kozma – Bognar from Biological Research Centre, Szeged, Hungary.

Light emission is measured in terms of Relative Light Units (RLU) per second and this corresponds to the rate of the light emission reaction for the colony [119].

The parameter RLU is a conversion factor that translates the number of moles of luciferin reacted into the RLU measurement by the instrument. This also accounts for the discrepancies in colony sizes (Fig 2.3), growth rate, and instrument characteristics.

'Diffusion' part of the model:

To account for the initial difference in cell density for each yeast patch that affects luminescence intensity (Fig 2.3) and for the non-uniformity of the solid media, the initial conditions were set for each patch (each experiment) individually, so that each experimental curve is considered for two regimes, "diffusion" and "phytochrome". The former starts from the application of luciferin with initially decreasing luminescence, which approaches an approximate steady state after 17-18 hours. The latter begins from the light treatment and continues to the end (Fig 2.5, A). The first regime fits separately to the diffusion part of the model and provides the initial luciferin level at the time of light application. Such decomposition of the initial conditions is introduced to describe the temporally changing substrate availability that emerges from the solid culture conditions. Modelling the solid media allows a much wider variety of experimental applications in conditions that are most common in yeast and synthetic biology.

Our experimental setup [112] involves the application of a relatively small amount of luciferin substrate (20 µl) to a yeast patch, growing in a 100-mm diameter plate on an agar gel of 5-7 mm thickness. We assumed that the initial decrease in luminescence level just after luciferin application predominantly resulted from the diffusion of substrate through the gel. This was confirmed by an additional experiment (Fig 2.4). The diffusion rate estimated from this experiment has a range of 3 to 10 mm/h, which is in good agreement with literature data for diffusion rate of small molecules in agar gel [120].

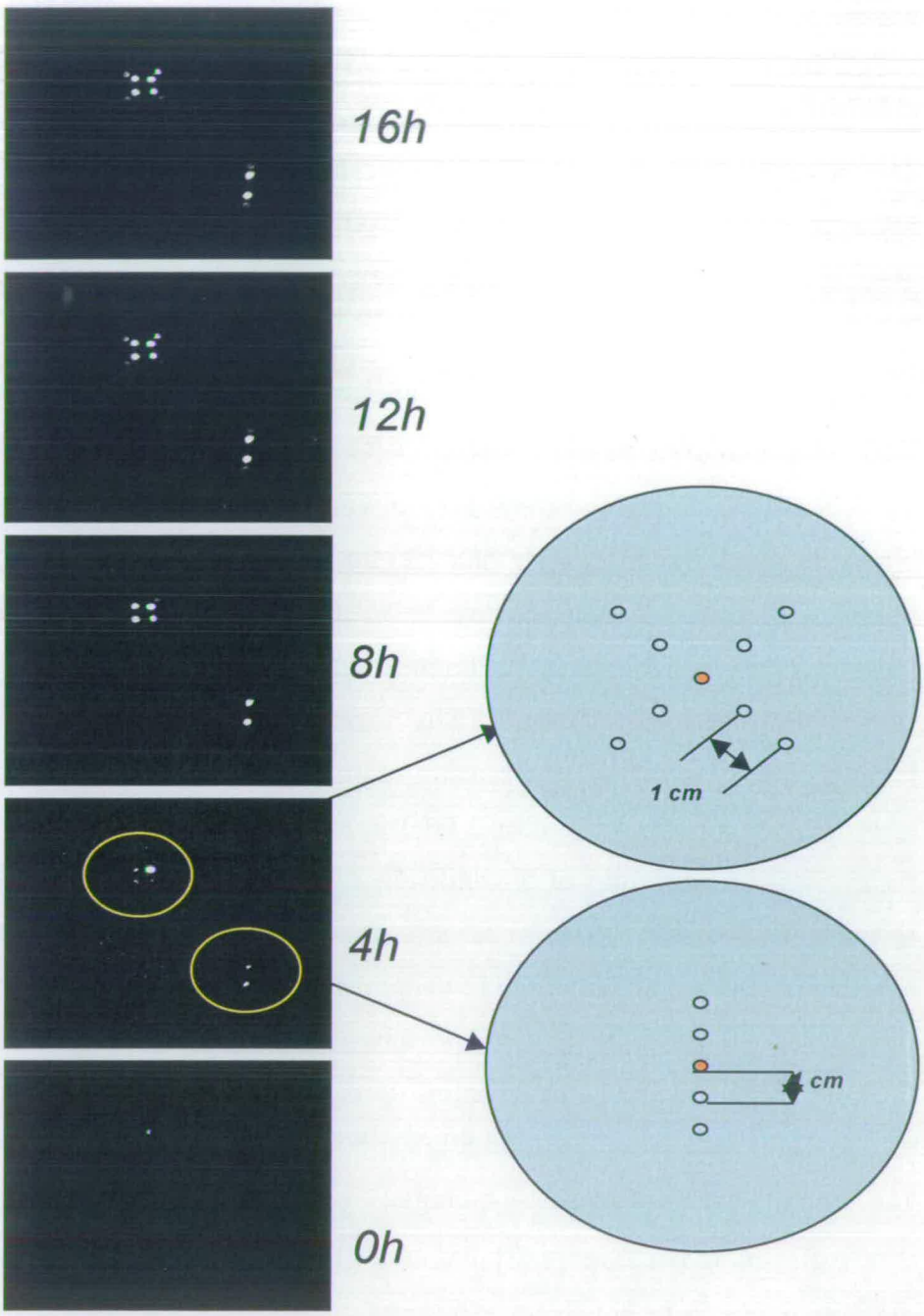


Figure 2.4 The experiment for the estimation of the diffusion rate.
Yeast colonies were grown from 20µl of OD_{600} 0.6 cultures and placed at 1cm intervals as the pattern indicates in the figure above. Luciferin was applied to the central well (orange on figure) and images were taken at 6min intervals for 26 hours. The rate of diffusion was calculated from the successive time intervals between colonies showing luminescence. The experiments have been performed by Oxana Sorokina and Laura E. Dixon, The University of Edinburgh.

Taking into account that the thickness of the gel is much smaller than the diameter of the plate, we assumed that the diffusion of luciferin could be described with the diffusion equation in polar cylindrical coordinates:

$$\frac{\partial S}{\partial t} = D\left(\frac{\partial^2 S}{\partial r^2} + \frac{1}{r} * \frac{\partial S}{\partial r}\right),$$

The particular solution of form

$$S = 2500000 * A + \frac{B}{time + 1} * e^{-\frac{r0^2}{4D*(time+1)}} \quad \textbf{(2.9)}$$

was found to fit experimental data with the best accuracy; $time + 1$ term is used to avoid singularity at time point zero. Here, S is the cytosolic luciferin concentration, D is the diffusion coefficient, r0 is the effective colony radius, A and B are constants of integration. Large numerical coefficient is added to take into account initial concentration of luciferine (2.5 mM) applied to the system. The value for diffusion coefficient D is individual for each experiment and depends on the agar density. The average value for D obtained from the fitting (see below) is 3×10^{-6} m^2/h, which is the same order of magnitude as the diffusion coefficient of similar compounds in agar gel provided by literature [121-123].

2.1. Fitting to experimental results

Parameter values for model equations were obtained from fitting the model to all the timeseries data from luciferase imaging (Fig 2.5 A, B), within the paramer ranges derived from the literature (see Table A.2). The parameter data for luciferase protein degradation rate was supported by additional experiments using cycloheximide treatment of yeast cultures constitutively expressing the luciferase gene (data not shown). The degradation rate constant is estimated to be 0.2-0.8 h^{-1}, which corresponds to a 0.8-3 hour half-life.

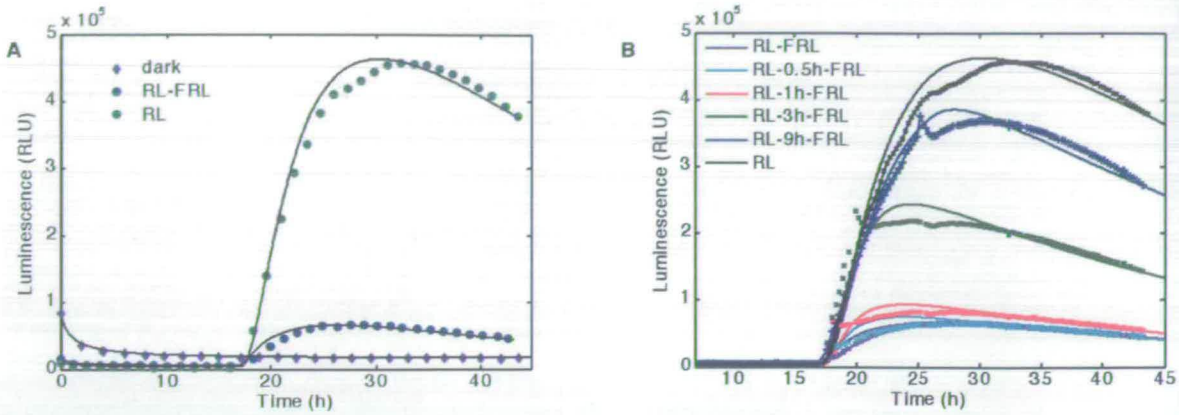


Figure 2.5 Time course of luciferase luminescence intensity in different light conditions

Experiment results (markers) and simulation (solid lines) are presented at the same scale. (A) Yeast cells containing PhyA_GBD and FHL_GAD were incubated in the dark. Luciferin was added at time zero with application of: no pulse (dark),10 min red pulse (RL), or 10 min red pulse followed immediately by 10 min far red pulse (FRL) at 18 h. (B) On-Off experiment with different time intervals between red and far-red pulses. Yeast cells containing both PHYA- GBD andd FHL-GAD were preincubated with PCB in the dark. Luciferin was added at time zero. At time 18 h all the cultures were given the 10 min red pulse (RL) or 10 min red pulse with subsequent 10 min far- red pulse at 0 (RL-FRL), 0.5 (RL-0.5h-FRL), 1 (RL-1h-FRL), 3 (RL-3h-FRL) and 9 h (RL-9h-FRL) after red pulse. A representative dataset of 5 replicates is shown.

This is similar to the value measured in yeast and mammalian cell cultures [119, 124] and plants [125].

The model was developed in SBTOOLBOX2 for MATLAB [126] and fitted with a particle-swarm optimisation algorithm [127] from the SBPD package in SBTOOLBOX2. Experiments were designed to cover all possible states of the system that have to be addressed in the model. We started with fitting the model to the simple experimental protocol, including dark conditions and red light application with or without the subsequent immediate far-red application (Fig 2.5, A). Dark experiments taken separately provided us with parameter values for the luciferase system (see Table A.1), namely the degradation and translation rates, that were fixed during the following optimization procedure. Light response parameter values were estimated from R and R-FR experiments. For that the model was simultaneously fitted to five sets of ON-OFF experiments, each containing seven experiments: R, dark and five combination

of R followed by FR with intervals 0h, 0.5h, 1h, 3h and 9h Fig 2.5, B). Thus, a total of 35 timeseries (each of 210-360 timepoints) were fitted simultaneously. Fitting results demonstrate a good accuracy (see Fig 2.5, A, B) with the root mean square deviation of $1.9*10^{-3}$. Spikes at the curves at certain time points correspond to the conditions of FRL application when plates have to be physically moved. They are not reproduced in the experiments and could be ignored while fitting.

As we aimed to account for increasing variability arising from solid culture conditions, our model parameters comprise some that are specific in each experiment. First of all, this relates to parameters corresponding to the 'diffusion' section (D, r0, A and B) as they establish the initial conditions by the time of light treatment. That parameters was obtained by fitting initial part of experimental curve to diffusion submodel by MATLAB fmins function. Secondly, parameter RLU that accounts for variability in colony size and growth rate also has to be locally estimated for each experiment. Therefore, besides the parameters identical for all experiments shown in Table A.1, the system contains 5 local parameters, namely RLU, D, r0, A and B, four of them (D, r0, A and B) were fitted to experimental data priot to main optimization run and the last one (RLU) have to been fitted in SBPD individually for each timeseries to obtain the results in the same scale.

Identifiability analysis

The MOTA tool was used to assess the identifiability of the model and reduce the number of free parameters in the fitting procedure [128]. We started from non- identifiable system with 17 parameters and followed the procedure described in [128]. After we fixed in series the functionally related parameters, we ended up with an identifiable 10-parameter system (Table A.2).

- 500 parameter sets were obtained from independent fittings from 500 random initial points
- Parameter sets were analysed to find statistical dependence between parameters

- Seven parameters were fixed to make Pool_Phy, k_base, m_luc1, K_rev, g_Luc, K_dis, k_kat
- The Final model contained 10 free parameters.

2.2. Model predictions

The refined model both captures qualitative dynamics and enables a quantitative description of the light switching behaviour. Moreover, it enables us to deduce which parameters would be critical for particular behaviours of the system.

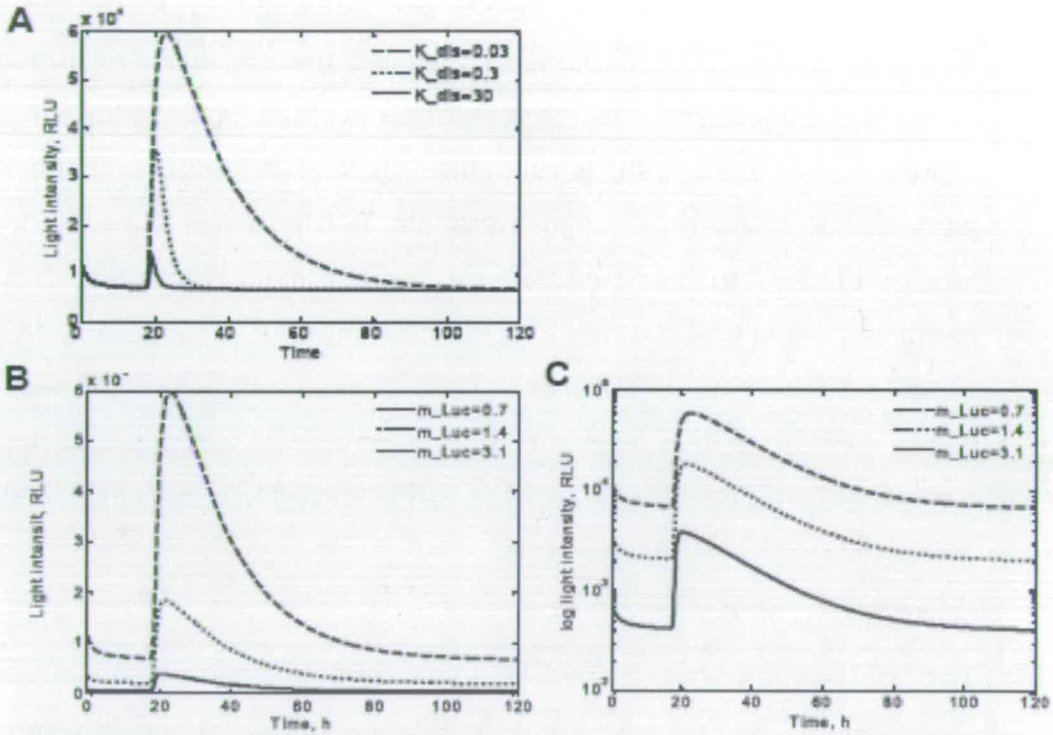


Figure 2.6 Investigation of the crucial parameters by simulation of the model
A. Simulation of different stability of Pr_FHL intermediate state. Dashed line corresponds to initial value of dissociation constant K_dis (h^{-1}), derived from fitting to the experimental curves. Dotted line corresponds to 10 times faster Pr_FHL complex dissociation. Solid line corresponds to 100-times faster complex dissociation.
B. Simulation of different stability of Luc reporter. Dashed line corresponds to initial value of degradation rate constant m_luc (h^{-1}), derived from fitting to experimental data. Dotted line corresponds to twice faster degradation rate of luciferase. Solid line corresponds to more than 4- times faster degradation rate.
C. Same as B, but in semilog scale. Reporter instability affects the amplitude of the signal but not the overall timing of events

By varying these parameters we showed that the predicted intermediate state during the photoconversion of Pfr_FHL is crucial to match the slow switching off of the observed LUC expression (Fig 2.6, A). The biochemical nature of this state as well as its experimental measurement is the subject of further experiments. Also, according to the model simulation (Fig 2.6, B, C), shortening of reporter protein half-life does not affect the longevity of reversal of the transcription activation but significantly reduces the intensity of the luminescence.

The light switch model also gives several predictions about the long-term system behaviour (Fig 2.7, A). In particular, we predicted based on experimental data for 50h, that in the experimental conditions considered the complete removal of the transcriptional activation effect should take a relatively long time (100 hours), and this has been confirmed by experiments (data not shown). Furthermore, with the given dynamics, we can manipulate subsequent applications of R and FR, to achieve a wide range of desirable profiles of transcription activation (Fig 2.7, B, C, and D). Fig 2.7, B illustrates the different types of behaviour of light input, which depend on the interval between R and FR treatments. It is clear from simulations that a small interval (2 min) between R and FR causes the increase of transcription rate and, accordingly, the increase of the luminescence intensity with time. Meanwhile, a longer interval (5 hours) produces the stable base line of input oscillations. On the basis of these simulations one can create a specific protocol of light input, combining the given modes as required, and thus obtaining the "square" shape (Fig 2.7, D) with two modes of light regime, and "sigmoid" shape (Fig 2.7, C) with three modes. Thus, the overall system could be used as a tool for the design of experiments with flexible perturbations of the system, for example, by changing the time intervals between.

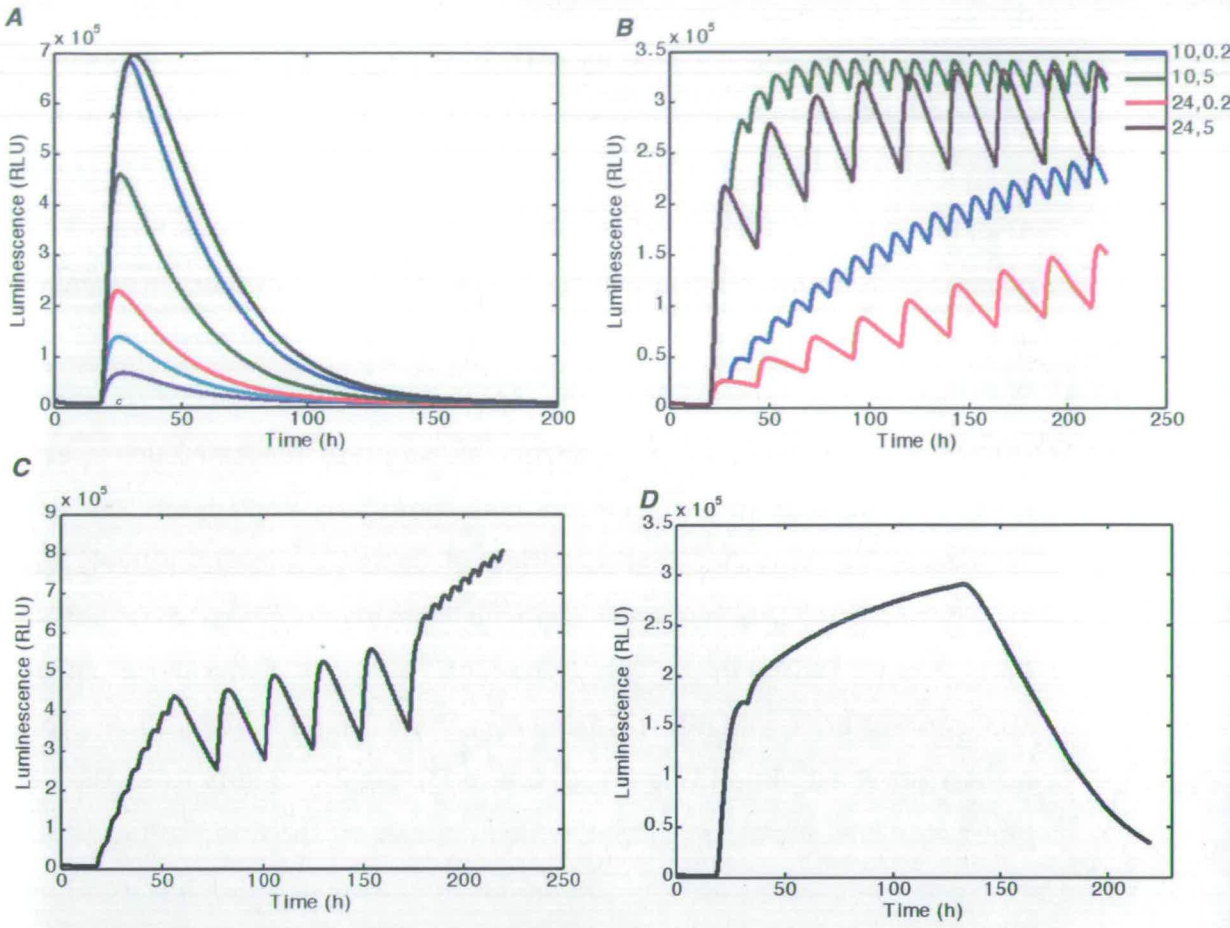


Figure 2.7 Model simulation and predictions

(A) Single peaks. Long term On-Off model simulation with far red light application in different time intervals after red pulse (0h, 0.5h, 1h, 3h, 9h, RL). Conditions are as in vivo On-OFF experiment but the time of simulation is prolonged until the complete removal of the signal, which corresponds to about 150h after ight treatment.

(B) Multiple peaks. Model simulation with different time intervals between red and far-red pulses. The green curve corresponds to 10 h between subsequent RL pulses and 5 h between neighboring RL and FRL pulses, the black curve to 24 h between RL pulses and 5 h between RL and FRL pulses, the blue curve to 10 h between RL pulses and 12 min between RL and FRK pulses, and the red curve to 24 h between RL pulses and 12 min between RL and FRL pulses.

(C) Possible profile of transcription activation with RL and FRL pulses of different frequencies of RL (RL-FRL is 2 min everywhere):5h-24h-5h- "sigma" shape.

(D) Possible profile of transcription activation with RL pulse at time 18 h and subsequent RL- FRL pulses with 2 h between RL and 2 min between RL-FRL until time 133 h- "square" shape.

2.3. Discussion

We developed a mathematical model that describes the system and fits the experimental data with great accuracy. The model incorporated experimental variability arising from the cultures on solid media, *via* substrate diffusion that corresponds to observations and sets the initial substrate concentration. To our knowledge, this is the first attempt to model the specificity of substrate distribution in solid cell culture. The model results demonstrate good overall accuracy (see Fig 2.5, A, B); nevertheless, the success of the fit varies between experiments. For example, in the On-Off experiments (Fig 2.5, B), the model fits better to the longer intervals (at least 1 hour between R and FR treatments), compared with the shorter treatment intervals, when FR is given immediately or 30 min after R. It can be seen from the time series that there is only a small quantitative difference between immediate and 30-min-delayed FR, while after a 1 hour delay the shape of the response resembles that of a single R treatment (Fig 2.5, B), differing only in amplitude. This qualitative shift in the system behaviour requires further analysis, which may shed light on mechanism of transcription activation by R light and inactivation by FR light in the system.

Our system does not display an instantaneous shutting off of target gene expression. It takes a substantial period of time to completely remove the Luc signal after FR treatment. Modelling suggests that this is not simply due to stability of the Luc reporter (see Fig 2.6, B, C), but rather reflects persistent PhyA activity. It should be noted that far-red exposure does not convert all the active PhyA into the Pr form, but by itself produces about 3% of Pfr form [116]. Additionally, we propose a residual physical interaction between the Pr form of PhyA and FHY1/FHL as a possible explanation for the slow kinetics. This was supported with model simulations that correspond to the experimental kinetics (Fig 2.6, A). However, the properties of the intermediate state remain to be determined.

Chapter 3. Modelling of the synthetic core oscillator

An understanding of the design principles of simple circuits would help in gaining an insight into more complicated system dynamics and provide the ability to reproduce and modify the rhythmic behaviour of the clock in a living organism.

The aim of the initial model development for the synthetic circuit is to determine whether a system comprising given biological components could establish oscillatory behaviour in a biologically plausible parameter range. Using the known kinetic characteristics of those components we could deduce which parameters would be crucial for the circuit design. We propose a phenomenological mechanistic ODE-based model that precedes the *in vivo* implementation of the network and is useful for a preliminary qualitative analysis of system behaviour. With known design and parameter constraints, the model would be able to suggest the optimal ways of improving the circuit composition and topology. After implementation of the system the model will be integrated with experimentally derived parameter values, which will enable a more realistic description of the circuit functionality to be given.

For our synthetic system we examined two possible circuit designs. One of them is based on the GAL regulon in yeast and is, actually, part of a more complicated circuit that has already been implemented in a yeast host [129] However, the original circuit has not been designed to be an oscillator, although potentially some useful kinetic information for the circuit components could be obtained from the publication [129]. The second design has been suggested by Laszlo Kozma- Bognar et al (personal communication) and is Tet-operon based. It has not been fully implemented *in vivo*, but it was intentionally designed as an oscillator and the kinetic parameters for its components are quite readily available in the literature.

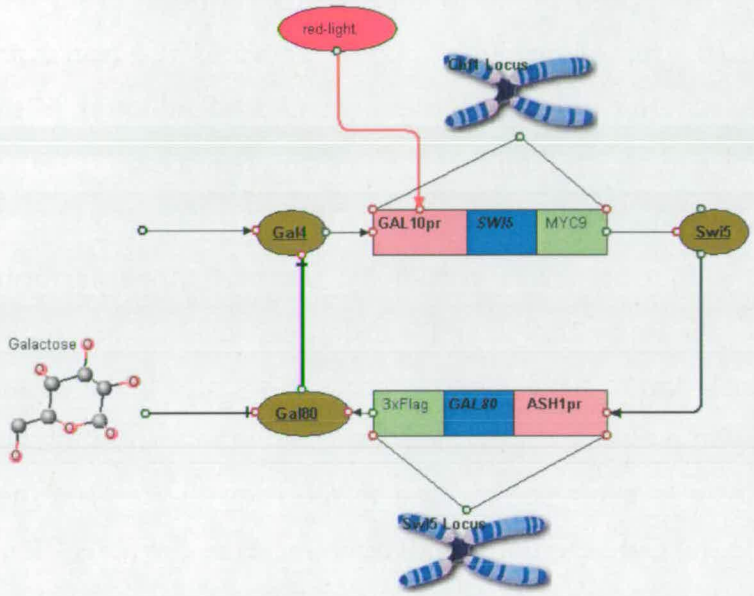


Figure 3.1 Small gal-operon based circuit outline.
All the genes are non-essential. Each red box designates a promoter, each blue box is the coding sequence of a gene and each green one denotes a tag sequence. Brown circules correspond to the proteins and the red circle correspond to the light input Solid lines represent regulation (arrow: upregulation, bar: inhibition) .

3.1. Gal-regulon based model.

The first network design is based on the yeast GAL regulon and represents part of the more complex system proposed in [129]. The GAL regulon is a natural yeast genetic system responsible for galactose metabolism. The interactions of its main components GAL4, GAL3 and GAL80 are widely described in the literature with particular attention given to the correspondent kinetic characteristics. Briefly, GAL4 is a transcription factor, which induces the expression of the set of galactose-responsive proteins. Two of them, GAL80 and GAL3, feed back in a negative and positive manner, respectively [130].

The proposed topology (Fig. 3.1) includes two gene elements, two mRNAs and three protein products, which comprise the single loop. The detailed structure of the interactions of model elements is represented at Figure 3.2. Here, GAL4 activates the transcription of another transcription factor, SWI5, placed under the GAL10 promoter, which, in turn, activates the expression of GAL80. The latter is nothing more than a known natural inhibitor of the GAL4 activator functions and, thus, completes the negative feedback.

We transformed the represented structure to a set of ODE. The model has to include sufficient delays crucial for the oscillatory performance, and so we consider the transcription and translation processes individually. As we assume the translation to be a cytosolic process, we also included the translocation of the protein into and out of the nucleus with the translation and degradation for the RNAs and proteins localised to the cytosol. We used the Hill functions to describe the transcriptional activation terms of the mRNA for SWI5 and GAL80. All the other reactions, including degradation, translocation, complex formation, and dissociation and translation were modelled with the mass action kinetics.

The model is based on the following assumptions:

1) We consider all the processes to take place in two compartments: the nucleus and cytosol, respectively

2) We consider the pool of GAL4 protein to be in a steady state, with no transcription and degradation explicitly included in the model

3) We consider the pool of galactose to be in a steady state

4) We consider mRNA transport from the nucleus lumped with the transcription rate.

5) We consider degradation of proteins and mRNA to be of first order and cytosol-localized.

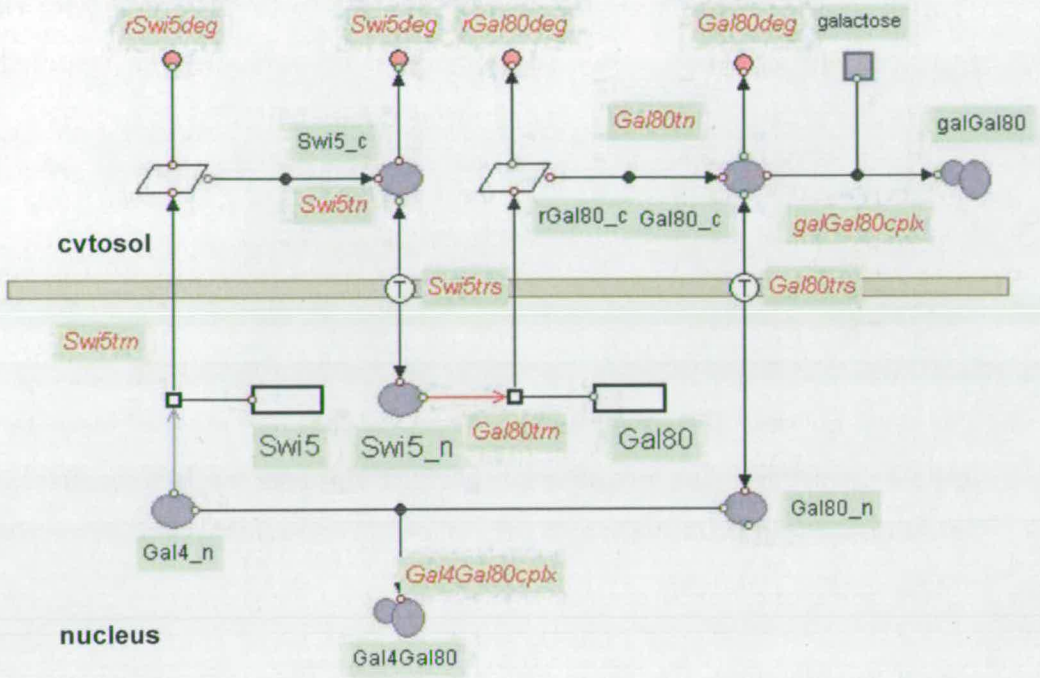


Figure 3.2 Detailed reaction structure of the gal- based model in SBGN format.

6) We suppose all the rate constants of mRNA translation to be the same and equal to 1 (1/h).

7) It is known, that GAL80 and GAL4 proteins participate in the process of activation and inhibition of transcription in their dimerized form. Instead of describing the dimerisation explicitly we increase the cooperativity for Swi5 activation, equating the corresponding Hill coefficient to 8, due to the dimer-binding 4 Upstream Activation Sequence (UAS) elements of the Gal10 promoter.

In an inducing medium (presence of galactose) GAL3 protein is activated and binds to GAL80. In the model we reduce this complexity by removing GAL3 and introducing the sequestration of GAL80 through its direct binding to the galactose molecule in the cytosol. This results in the release of free GAL4 capable of inducing transcription.

The nucleocytoplasmic shuttling of the repressor GAL80p is known to play a key role in the signal transduction process of the *GAL* genetic switch of *Saccharomyces cerevisiae* [131]. In non- inducing, non-repressing conditions, the expression of GAL genes is not detectable as GAL80, a shuttling repressor

protein, represses the transcription by binding to GAL4 bound with DNA. The translocation of GAL80 between the cytoplasm and nucleus is assumed to be a reversible process with the distribution coefficient of K_{nucDIS}. The value of K_{nucDIS} is equal to the ratio of the concentrations of free monomeric GAL80 in the cytoplasm to that in the nucleus [132].

$$K = \frac{[Gal80\_c]}{[Gal80\_n]}$$

The term for the SWI 5 protein transcription contains the '*lightresp*' term. This imitates if needed the introduction of light input to the system. In this model situation it could be only On and Off, which corresponds to unit or zero value for the parameter *lightresp*.

Hence, the reactions of the model have the following mathematical description:

$$galGal80_{cplx} = K1\_galGal80_{cplx} * [Gal80\_c] * [gal] \quad \textbf{(3.1)}$$

$$Gal80_{trl} = p\_gal8o * [rGal80\_c] \quad \textbf{(3.2)}$$

$$Gal4Gal80_{cplx} = K3\_Gal4Gal80_{cplx} * [Gal80\_n]^2 * [Gal4\_n]^2 \quad \textbf{(3.3)}$$

$$Gal4Gal80_{diss} = K\_Gal4Gal80_{diss} * [Gal4Gal80] * K3\_Gal4Gal80_{cplx} \quad \textbf{(3.4)}$$

$$galGal80_{diss} = K\_galGal80_{diss} * [galGal80] * K1\_galGal80_{cplx} \quad \textbf{(3.5)}$$

$$Swi5_{trn} = \frac{(n\_Swi5 * [Gal4\_n]^{a\_Swi5}) * (1 + lightresp * b\_Swi5)}{g\_Swi5^{a\_Swi5} + [Gal4\_n]^{a\_Swi5}} \quad \textbf{(3.6)}$$

$$Swi5_{trl} = p\_Swi5 * [rSwi5\_c] \quad \textbf{(3.7)}$$

$$Gal80_{deg} = m\_Gal80 * [Gal80\_c] \quad \textbf{(3.8)}$$

$$rGal80_{deg} = m\_rGal80 * [rGal80\_c] \quad \textbf{(3.9)}$$

$$rSwi5_{deg} = m\_rSwi5 * [rSwi5\_c] \quad \textbf{(3.10)}$$

$$Swi5_{deg} = m\_Swi5 * [Swi5\_c] \quad (3.11)$$

$$Gal80_{trs} = K5 * ([Gal80\_c] - KnucDIS * [Gal80\_n]) \quad (3.12)$$

$$Swi5_{trs} = K6 * [Swi5\_c] - K\_6 * [Swi5\_n] \quad (3.13)$$

$$Gal80_{trn} = \frac{n\_Gal80 * Swi5\_n^{a\_Gal80}}{g\_Gal80^{a\_Gal80} + [Swi5\_n]^{a\_Gal80}} \quad (3.14)$$

The model equations are:

$$\frac{d}{dt} galGal80 = galGal80_{cplx} - galGal80_{diss} \quad (3.15)$$

$$\frac{d}{dt} Gal4Gal80 = Gal4Gal80_{cplx} - Gal4Gal80_{diss} \quad (3.16)$$

$$\frac{d}{dt} Gal80\_c = galGal80_{diss} + Gal80_{trl} - galGal80_{cplx} - Gal80_{trs} - Gal80_{deg} \quad (3.17)$$

$$\frac{d}{dt} Gal80\_n = Gal80_{trs} - Gal4Gal80_{cplx} + Gal4Gal80_{diss} \quad (3.18)$$

$$\frac{d}{dt} rGal80 = Gal80_{trn} - rGal80_{deg} \quad (3.19)$$

$$\frac{d}{dt} rSwi5 = Swi5_{trn} - rSwi5_{deg} \quad (3.20)$$

$$\frac{d}{dt} Swi5\_n = Swi5_{trs} \quad (3.21)$$

$$\frac{d}{dt} Swi5\_c = Swi5_{trn} - Swi5_{trs} - Swi5_{deg} \quad (3.22)$$

Here, the reactions *galGal80cplx*, *Gal4Gal80cplx*, *galGal80diss* and *Gal4Gal80diss* correspond to the association and dissociation of the respective complexes. *Trn* reactions correspond to the transcription, *trl* to the translation and *deg* for the

degradation of the corresponding species. (See parameter descriptions in the TableB.1)

As we considered the pool of GAL4 protein (Pool 1) and galactose (Pool 2) to be constant, we operate with two conservation laws,

$$\begin{aligned} Gal4\_n &= -Gal4Gal80 + Pool\_1\_ \\ gal &= -galGal80 + Pool\_2\_ \end{aligned} \qquad \textbf{(3.23)}$$

Therefore, the overall model consists of 10 species with 8 independent species, 14 reactions, and 2 conservation laws with 27 parameters.

The parameter value ranges were derived from the literature, where available (Table B.1) The half-life and average mRNA concentration data were adopted from the Genome-Wide Expression Page [133] . Dissociation constants for GAL4GAL80 and GALGAL 80 complexes were taken from [132, 134]. There is no experimental evidence for the precise value of the Hill coefficient for GAL80 transcription activation by SWI5 protein, so it was set to 2 arbitrarily as a biologically reasonable value.

Switch –like behaviour of the galactose operon has been described in detail by [132, 135] with mathematical models suggesting the existence of two different steady states in the presence and absence of galactose. Moreover, Smidtas *et al* in 2006 proposed that the galactose-dependent system operates in response to the gradient of galactose and that there is only one steady state, which the system returns to after the gradient disappears [136].

Therefore, two possible types of system behaviour were expected, oscillatory and switch-like. The results of simulation demonstrate the switch-like behavior of the system, which depends on the presence of galactose. Figure 3.3 shows that in presence of galactose GAL4 protein dissociates from the complex with GAL80, so that its concentration increases. It activates the transcription of the Swi5 gene and the subsequent transcription of the Gal80 gene, so an increase in the concentrations of SWI5 and GAL80 mRNA and protein is observed.

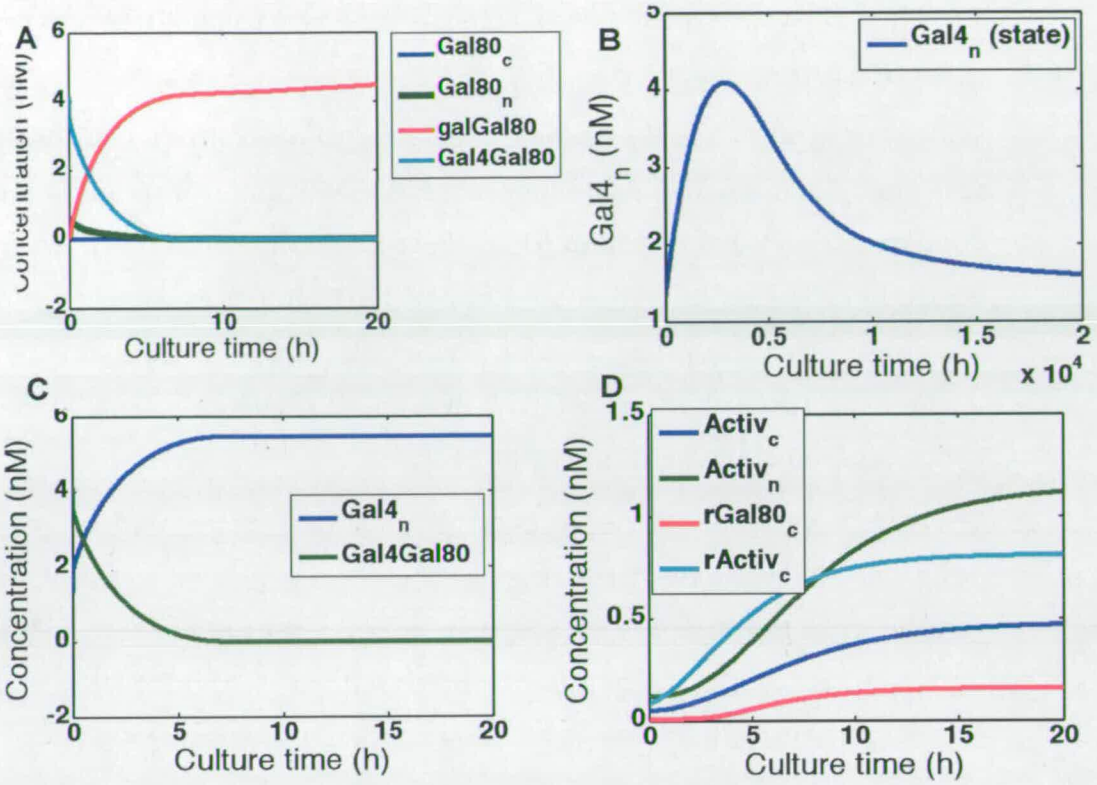


Figure 3.3 Results of simulation of the Gal-based model.

In the presence of galactose the release of the GAL4 protein from the complex with GAL80p induces the transcription of the activator gene, which in turn activates the transcription of GAL80. A. The release of GAL80 from GAL4GAL80 complex is accompanied with the increase in concentration of GALGAL80 complex B. Transient increase in GAL4 concentration caused by adding of galactose. C. The increase of concentration of free GAL4 results from its release from the complex with GAL80, which entails the decrease in the GAL4GAL80 amount. D. The appearance of the free GAL4 induces the expression of GAL80 and SWI5 ('activ' here) RNAs.

Synthetic circuit components in the above system are of yeast origin, which means inevitable cross-talk with the natural yeast cell metabolism.

As the initial mathematical model for the Gal4–based system predicted switch-like rather than oscillatory behaviour of the synthetic circuit, we proceeded with another design, presented in the following section

3.2. Tet-operon based model

The second design of the synthetic core oscillator is Tet-operon based. It has been proposed by Laszlo-Kozma-Bognar et al (personal communication). It represents a simple two-component feedback loop where the bacteria-derived

Tet-ON [137]system is used as an activator, and yeast Ssn6-Tup1 as a repressor [138] (Fig 3.4).

Widely used in synthetic biology, the tetracycline–controlled transcriptional activator, tTa, is a fusion protein between the TetR and transcription activation domain, a C-terminal region of VP16 (herpes simplex virus HSV). It stimulates transcription when bound to an array of tetO sequences fused to a minimal promoter derived from human cytomegalovirus. Tetracycline (Tc) or doxycycline (Dox) work as effector molecules, preventing DNA binding and, hence, abolishing transcription [139-142].

The system described above is an original Tet-OFF system and is useful when design of the synthetic circuit implies the occasional turning-off of an active gene. In cases in which it is necessary to keep the gene of interest inactive, with infrequent activation, it seems more appropriate to use the Tet-ON system, with contains tetracycline reverse transcriptional activator (rtTA), which corresponds to the complementary mutant system that reacts on the presence of the effector in the opposite fashion: Dox stimulates the transactivator's conformational change, which is needed for the binding to the promoter and activation of transcription

In our circuit design the gene for rtTA is placed under the yeast CUP1 promoter. CUP1 is a metallothionein protein involved in the yeast Cu ion detoxification system. Copper cooperatively binds to Ace1 protein, forming the tetra-Cu cluster, which is able to activate transient expression from the CUP1 promoter [143].

Ssn6 and Tup1 are conserved proteins, which promote the highly efficient repression of a wide spectrum of target genes in *S. cerevisiae*. The TuP1-Ssn6 complex consists of four Tup1 and one Ssn6 molecules. To abolish transcription it has to be recruited to the promoter by a corresponding sequence-specific DNA-binding protein. Several mechanisms for the repression have been proposed so far. Among them are the direct interference with the activator, altering the local chromatin structure by setting the repressive nucleosome positioning, and the breach of the general transcriptional machinery by

affecting the polymerases I and II complexes [144-146]. In our design the Ssn6-Tup1 complex specifically interacts with the LexA binding site, repressing the transcription from the CUP1 promoter and, thus, completing the feedback loop.

3.2.1. Model description and simulation

The model for the circuit described above is mechanistic and includes most of the phenomenological steps. In particular, we consider:

1) Two compartments: nucleus and cytosol

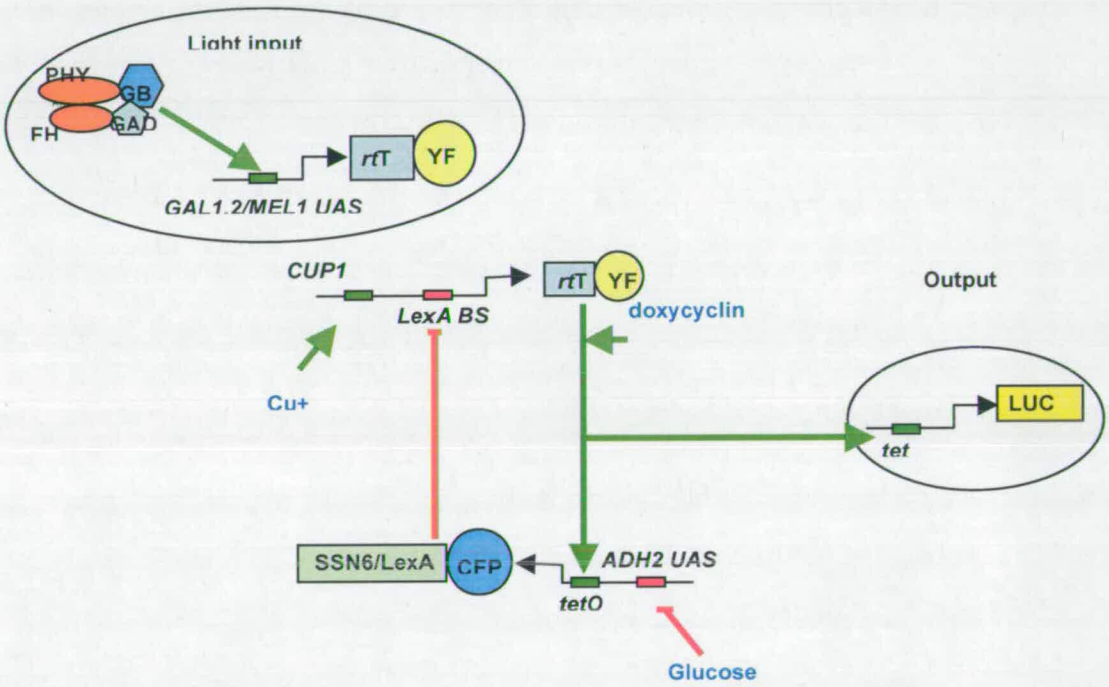


Figure 3.4 The design of the rtTA-based circuit
rtTA transcativator is placed under the CUP1 promoter and can be activated by adding $CuSO_4$ to the media. The expression of rtTA is inhibited by the SSN6/LexA complex directly bound to the LexA binding site. ADH2 UAS fused to the tetO promoter allows the fine regulation of the level of inhibitor expression. The light input could be provided to the system by means of the light switchable expression system described in Chapter 2. In this case the artificial PhyA-FHL system will determine the additional transcription for the rtTA components from the fused Gal1 promoter sequence.

2) A copper-Ace1 ternary cluster with distinct kinetics needed for the induction of transcription from the Cup1 promoter (reactions ***(3.24)*** and ***(3.25)***)

3) rtTA dimerization is introduced explicitly (reaction ***(3.28)***)

4) rtTA homodimer translocation to the nucleus where it binds two dox molecules one after another (reactions ***(3.29-3.31)***)

5) Ssn6 accounts for the Ssn6-Tup1 complex. Ssn6 protein is translocated to the nucleus where it represses the transcription of rtTA (reactions ***(3.37)***, ***(3.25)***)

6) All the reactions for degradation are of the first order and localized to both compartments (reactions ***(3.26)***, ***(3.32)***, ***(3.34)***, ***(3.36)*** and ***(3.38)***)

Hence, the reactions considered in the model are:

$$Ace1\_Cu_{cplx} = K\_Cu\_cplx * Ace1 * Cu^4 - K\_Cu\_diss * Ace1\_Cu \qquad (3.24)$$

$$rtTA_{trn} = \frac{n\_rtTA * Ace1\_Cu^{a\_rtTA}}{g\_rtTA^{a\_rtTA} + Ace1\_Cu^{a\_rtTA}} * \frac{1}{1 + \frac{Ssn6\_n^{a\_i\_Ssn6}}{Ki^{a\_i\_Ssn6}}} \qquad (3.25)$$

$$r\_rtTA_{deg} = m\_r\_rtTA * r\_rtTA \qquad (3.26)$$

$$rtTA_{trl} = p\_rtTA * r\_rtTA \qquad (3.27)$$

$$rtTA_{\dim} = K\dim\_cplx * rtTA\_c^2 - K\dim\_diss * rtTA\dim\_c \qquad (3.28)$$

$$rtTA_{trs} = K1 * rtTA\dim\_c - K2 * rtTA\dim\_n \qquad (3.29)$$

$$rtTA\dim\_dox_{cplx} = Kdox\_cplx * rtTA\dim\_n * dox - Kdox\_diss * rtTA\dim\_dox\_n \qquad (3.30)$$

$$rtTA\dim\_2dox_{cplx} = Kdox\_cplx * rtTA\dim\_dox\_n * dox - Kdox\_diss * rtTA\dim\_2dox\_n \qquad (3.31)$$

$$rtTA\dim\_n_{deg} = m\_rttA\dim\_n * rtTA\dim\_n \qquad (3.32)$$

$$Ssn6_{trn} = \frac{n\_Ssn6 * rtTA\dim\_2dox\_n^{a\_Ssn6}}{g\_Ssn6^{a\_Ssn6} + rtTA\dim\_2dox\_n^{a\_Ssn6}} \qquad (3.33)$$

$$r\_Ssn6_{deg} = m\_r\_Ssn6 * r\_Ssn6 \qquad (3.34)$$

$$Ssn6_{trl} = p\_Ssn6 * r\_Ssn6 \qquad (3.35)$$

$$Ssn6_{trs} = K3 * Ssn6\_c - K4 * Ssn6\_n \quad (3.36)$$

$$Ssn6\_n_{\deg} = m\_Ssn6\_n * Ssn6\_n \quad (3.37)$$

Here, *Ace1_Cu* is the copper-Ace1 ternary clasters, *r_Ssn6* and *r_rtTA* are the RNAs, *rtTa* and *Ssn6* ar the corresponding proteins which may have nuclear or cytosolic location (*_n* or *_c*). *rtTAdim*, *rtTAdim_dox* and *rtTadim_2dox* correspond to the dimer that could be associated with one or two dox molecules. Reaction subscripts *trn*, *trl*, *trs*, *dim*, *cplx* correspond to the transcription, translation, translocation, dimerization and association, respectively. Parameters are defined in Table B.2.

The ODE's are presented below:

$$\frac{d}{dt} Ace1\_Cu = Ace1\_Cu_{cplx} \quad (3.38)$$

$$\frac{d}{dt} Ssn6\_c = Ssn6_{trl} - Ssn6\_c_{\deg} - Ssn6\_trs \quad (3.39)$$

$$\frac{d}{dt} Ssn6\_n = Ssn6_{trs} - Ssn6\_n_{\deg} \quad (3.40)$$

$$\frac{d}{dt} r\_Ssn6 = Ssn6_{trn} - r\_Ssn6_{\deg} \quad (3.41)$$

$$\frac{d}{dt} r\_rtTA = rtTA_{trn} - r\_rTA_{\deg} \quad (3.42)$$

$$\frac{d}{dt} rtTA\_c = rtTA_{trl} - 2 * rtTA_{\dim} \quad (3.43)$$

$$\frac{d}{dt} rtTA\dim\_2dox\_n = rtTA\dim\_2dox_{cplx} \quad (3.44)$$

$$\frac{d}{dt} rtTA\dim\_c = rtTA_{\dim} - rtTA_{trs} \quad (3.45)$$

$$\frac{d}{dt} rtTA\dim\_c = rtTA_{\dim} - rtTA_{trs} \quad (3.46)$$

$$\frac{d}{dt} rtTA\dim\_n = rtTA_{trs} - rtTA\dim\_dox_{cplx} - rtTA\dim_{\deg} \quad (3.47)$$

Mass action kinetics were used to describe the degradation of proteins and RNAs, translocation of proteins between the nucleus and cytosol, complex

formation and dissociation and translation of mRNA to cytosolic proteins. Hill functions were used for the transcriptional activation and inhibition terms of the mRNA for rtTA and Ssn6.

The parameter values for the models were obtained from the literature [143, 147-151] (Table B.2). Several groups have performed the kinetic studies for the interactions of TetR repressor and tetracycline, and the resulting constants vary in a range of four orders of magnitude. For instance, the association constant value for TetR-[Tc-Mg], $K_a = 3*10^9 M^{-1}$, has been determined by [148]. The value for the same constant derived by [147] varies from $7*10^4 M^{-1}$ to $9.0*10^6 M^{-1}$, depending on the measurement protocol. We therefore adopted the mean values for the available constants and allowed them to vary in a range of 2 orders of magnitude. Rates of degradation were taken from [133], where possible. The resulting list of initial parameter values with allowed ranges is presented in Table B.2.

We also considered another possible mechanism for dox- rtTA association where two molecules of dox bind the rtTA homodimer simultaneously. In this case the reactions ***(3.30)*** and ***(3.31)*** become a single reaction of the form:

$$rtTA\dim\_2dox_{cplx} = Kdox\_cplx * rtTA\dim\_n * dox^2 - (Kdox\_diss * rtTA\dim\_2dox\_n) \quad \textbf{(3.48)}$$

Using the quasi-steady state assumption, we derived the relationship between the parameters that results in a situation where more complicated sequential dox-binding could be approximated by a more compact simultaneous one (Appendix C). The final relationship has a form:

$$k_{-1}^{II} \;>>\; k_2^{II}[B], \quad \textbf{(3.49)}$$

where k_{-1}^{II} is first dox dissociation rate constant; k_2^{II} – second dox association rate constant and [B] corresponds to the dox concentration

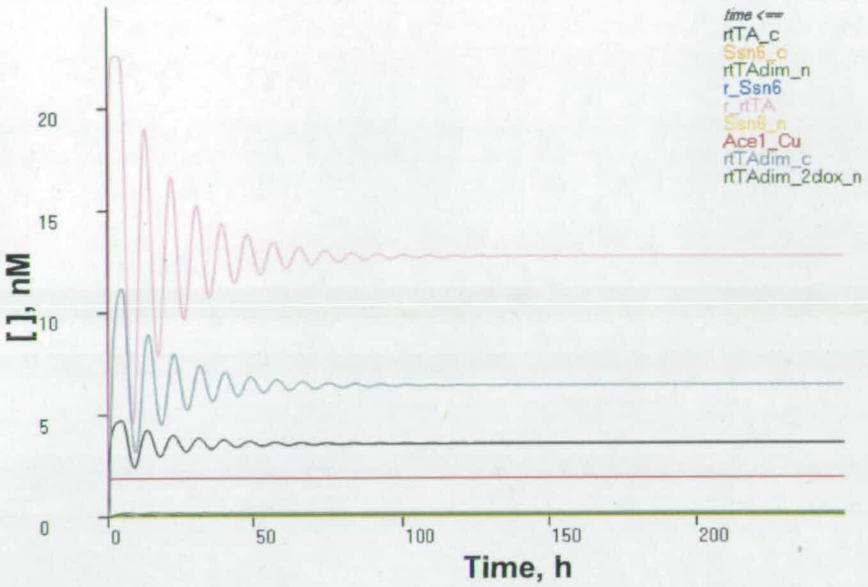


Figure 3.5 The results of simulation of the 3orderModel
Damping oscillations with the 8h period are shown

This means that the second-order mechanism could be approximated by the third-order one only in a situation where the rate of dissociation of the single dox complex is much faster than the rate of the second dox association (Appendix C).

For the parameter values presented in Table B.2 this inequality hardly works, although we have a very little idea about the real parameter values and, especially, whether the rate for the second dox binding would be the same as for the first one.

Nevertheless, until we have got the first experimental results, we are considering both cases separately, which allows us to explore the parameter space rigorously.

As was expected, the two proposed mechanisms resulted in different system behaviours. The results of simulations with Oscill8 software [152] are presented at Figure 3.5. The model with simultaneous dox binding (3orderModel) demonstrates damping in the first 50 hours. The period of damping oscillations is about 8 hours. The concentrations for rtTA and Ssn6 fall in a different range and could hardly be seen on the same graph.

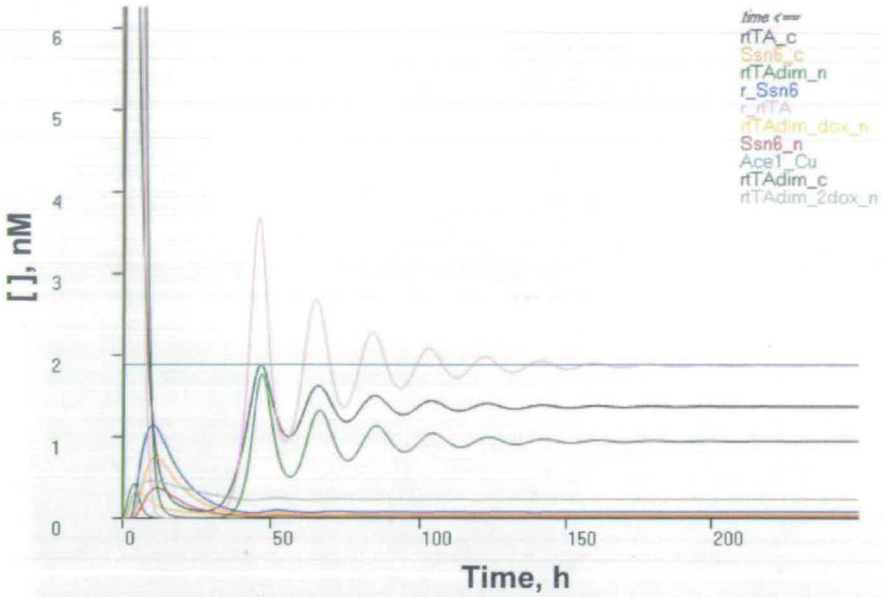


Figure 3.6 The results of the simulation of the model for the successive dox-binding (2orderModel).

Damping oscillations with an 18h period spring up after an initial large peak, followed by a delay about 20h.

The model with non-simultaneous dox binding (2orderModel) demonstrates slightly different dynamics (Fig 3.6). Although we are using the same numerical value for the rate of dox-rtTAdim association (1.189), this constant will have different units: $nM^{-1}h^{-1}$ for the 2orderModel and $nM^{-2}h^{-1}$ for the 3orderModel.

At the same dox concentration (225 nM) 3orderModel has an advantage in association rate of two orders of magnitude. This results in more rapid accumulation of the activatory complex.

The presence of the first high-amplitude peak in both cases stems from the initial accumulation of rtTA after the introduction of copper into the system. Nevertheless, the distinct mechanisms proposed in the two models account for the difference in subsequent dynamics.

Whereas in the 3orderModel the active rtTAdim2dox complex appears rapidly and is able to induce the Ssn6 transcription during the first 10 hours, the second system needs time for accumulation of a quantity of rtTAdim2dox sufficient for Ssn6 induction, so that initial peak is followed by a 20 hour delay before the system starts oscillating with a period of about 18 hours. The difference in

period length between the two models is also governed by the rate of active rtTA-dox complex accumulation. The model with simultaneous dox has an advantage in this rate about an order of magnitude, which results in smaller period and amplitude.

3.2.2. The investigation of the model parameter space with the bifurcation analysis

As we are looking for sustainable oscillations, it is necessary to either tune the parameter values in such a way as to bring our system to the oscillatory region or modify the structure of the system so that it is able to perform the desired behaviour with the given parameter values.

At the stage when we are primarily interested in the qualitative features of the system's dynamical behaviour, very useful information could be obtained from the dynamical constraints derived from the bifurcation structure. Bifurcation analysis allows the determination of how the system performance depends on the parameter values.

As a short introduction, if we define the system of ODE as $x' = f(x,p)$, $x \in R^n, p \in R^m$, where x is a concentration vector and p is parameter vector, we can suppose a smooth curve $p = v(\alpha): R \to R^m$ through the parameter space such as $v' \neq 0$. We are looking for the steady-state solution $\hat{x}(\alpha)$, so that $f(\hat{x}(\alpha), p(\alpha)) = 0$.

The idea of the nature of the steady state could be deduced from the eigenvalues, $\lambda_i(\alpha)$, of the Jacobian matrix, evaluated in the respective steady state:

$$A = \frac{\partial f}{\partial x}(\hat{x}(\alpha))$$

The solution $\hat{x}(\alpha)$ is stable if $\mathrm{Re}\ \lambda_i(\alpha) < 0$ for all i, meaning that all the trajectories in the vicinity of $\hat{x}(\alpha)$ tend to $\hat{x}(\alpha)$. Otherwise it is unstable, that is all the nearby trajectories tend away from $\hat{x}(\alpha)$. A local bifurcation occurs in the

system at $p^* = v(\alpha_0)$ as α passes through a value, when for some eigenvalue λ_k, $\mathrm{Re}\,\lambda_k(\alpha_0) = 0$.

In biological systems, the most important types of bifurcations are the saddle-node (SN), saddle-node on invariant circuits (SNIC) and Hopf (HB) bifurcations [153]. The latter is a hallmark of negative feedback and is known as a classic way of generating spontaneous oscillations in a biochemical system [153]. In the case of HB the complex conjugate pair of eigenvalues should satisfy three conditions:

1) $\mathrm{Re}\,\lambda_k(\alpha_0) = 0$
2) $\frac{\partial \mathrm{Re}\,\lambda}{\partial \alpha}(\alpha = 0) \neq 0$
3) $\mathrm{Im}\,\lambda \neq 0$

In our work we used Oscill8: a suite of tools for analysing large systems of ODE, deliberately developed for the bifurcation exploration of the dynamical systems and circadian models in particular [152]. Oscill8 works with the SBML and XPP format of ODE model and provides a number of very useful facilities, such as period/min/max sampling, bifurcation search/match/optimization and many others [153]. We primarily used the option of generating one-and two parameter bifurcation diagrams to find the region of HB. For the analysis we have chosen the model describing the simultaneous dox binding mechanism, as it performs better from the point of view of the resulting oscillations (higher amplitude, absence of the initial 20h lag).

It is necessary firstly to find the regions of oscillations. Given that negative feedback typically produces oscillations via Hopf bifurcation, which could happen due to the time delay of the loop, we have to focus on the parameters that provide this effect, namely parameters for degradation, complex association/dissociation and translocation. Given the initial settings from Table B.2, we generated the steady state one-parameter continuations for each parameter from the set to obtain the ordered collection of bifurcation points.

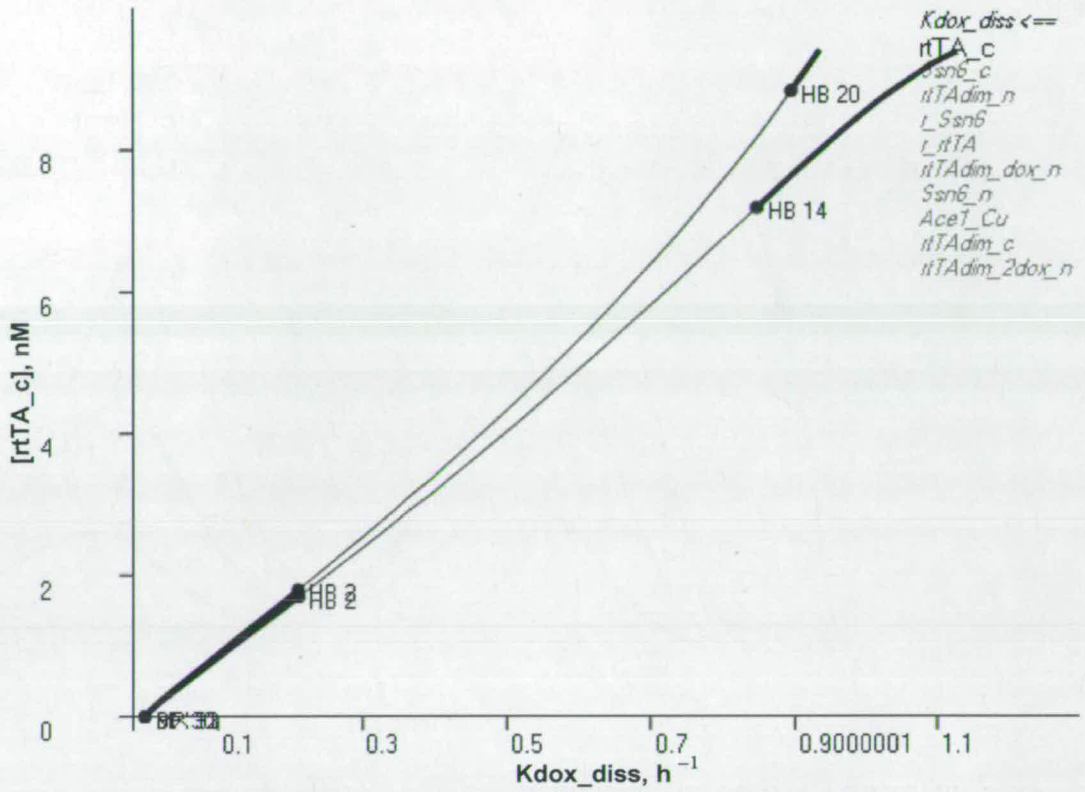


Figure 3.7 The one-parameter continuation for the rate of the dissociation of the dox-rtTA complex Kdox_diss.

Here, HB2 is a supercritical Hopf bifurcation. The system performs sustained oscillations within the range for Kdox_diss from 0.2 to 0.8 h^{-1} (right curve). The left curve (HB20) corresponds to the model with the introduced term for autoactivation (see the subsection 3.2.3).

The order of the bifurcations presented on the one–parameter bifurcation diagrams is defined by listing the encountered bifurcation points while tracing out the steady state curve from the minimum allowed parameter value to the maximum. The resultant diagrams are used to identify the particular parameter whose value could be most easily tuned to attain the region of Hopf bifurcation.

Fortunately, with a successful initial attempt, we localized a HB from the one-parameter continuation for the rate of dissociation of the dox-rtTA complex at the value K_{dox_diss}= 0.2 h^{-1}, which means that system, oscillates within the domain [0.2, 0.8] h^{-1} for K_{dox_diss} (Fig 3.7). The values from this range satisfy the allowed literature constraints (Table B.2)

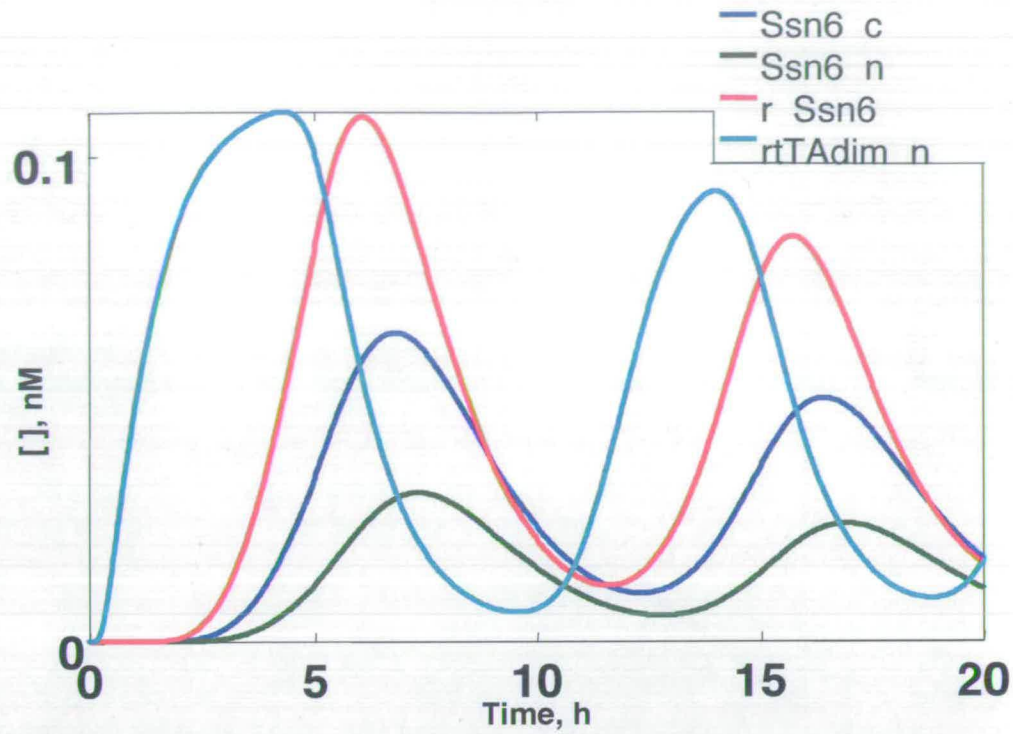


Figure 3.8 The time series simulation of the model.
The model was simulated with the parameter for the dissociation rate of dox-rtTA adjusted to 0.7 h^{-1}. The system oscillates with a period of about 10h.

Moving the value for K_{dox_diss} from 0.07 to 0.7 h^{-1} we obtained a system that produces sustainable oscillations with a period of about 10h (Fig 3.8).

3.2.3. Adding the autoactivation component

Another way to bring the system to the desired behaviour was to change its design in such a way that the resulting topology could successfully perform the oscillations within the initial parameter ranges. Therefore, we proceeded with the structural changes.

One possible way to reach the robust sustainable oscillatory behaviour was to introduce autoregulation for one of the components; for instance the activator term. The combination of negative feedback with a positive autoregulation is supposed to provide an overall increase in robustness and the ideal conditions for evolving the oscillations [17]. This also gives a physiological advantage as it endows the oscillatory system with the ability to tune the frequency while

retaining the amplitude, hence, similar topological structures are widely described among the natural oscillators [17].

From the engineering point of view, we may fuse the tetO sequences to the promoter, which controls the expression of rtTA, thus, providing the autoactivation for the rtTA expression. Varying the number of these sequences we can modulate the strength of the autoactivation. Thus we could have the activator inducing its own expression as well as the expression of the repressor.

To investigate the potential capacity of the new design to produce the sustained oscillations we changed the model for the synthetic network presented in the previous section in the following way. We added the additional Hill term for the self-activation to the equation for transcription of rtTA, which gained a form of:

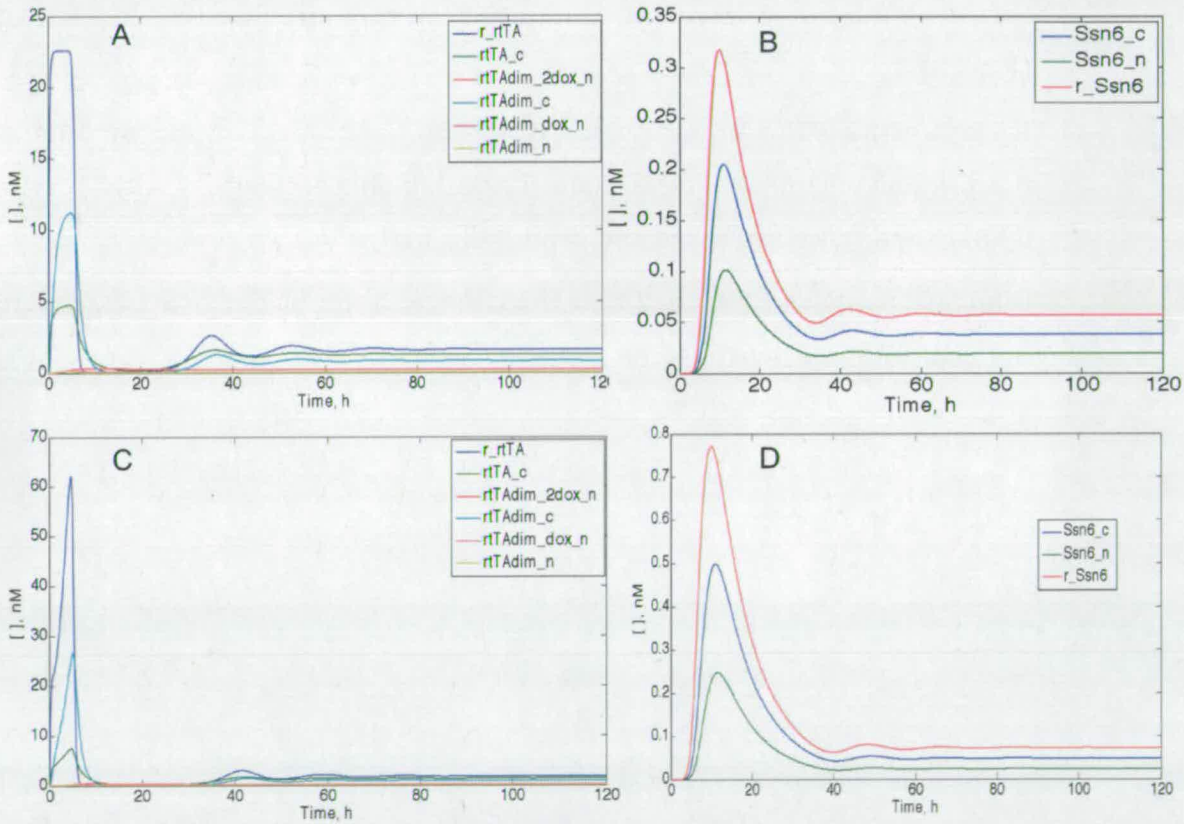


Figure 3.9 The results of simulation of the models for the rtTA based synthetic oscillator with successive Dox-binding mechanism, with an added term for the positive self regulation of the expression of the activator component.

Upper figures corespond to the initial model behaviour. Lower figures correspond to the system with added autoactivation . Note the increase in the component concentrations for the lower figures (axis Y) A. Plotted is a time series for the rtTA component.The initial peak is followed by subsequent low ampitude damping oscillations. B. Time series for the inhibitor (Ssn6) component. C. Self –activated rtTA demonstrates a very similar performance to the non- activated one, although the amplitude of the initial peak is 3 times higher. D. The same is true for Ssn6.

$$rtTA_{trn} = \frac{n\_rtTA * Ace1\_Cu^{a\_rtTA}}{g\_rtTA^{a\_rtTA} + Ace1\_Cu^{a\_rtTA}} * \frac{1}{1 + \frac{Ssn6\_n^{a\_i\_Ssn6}}{Ki^{a\_i\_Ssn6}}} *$$

$$(1 + \frac{rtTA\dim\_2dox\_n^{a\_aut}}{g\_Ssn6^{a\_aut} + rtTA\dim\_2dox^{a\_aut}}) \qquad \textit{(3.51)}$$

where a^{aut} is the Hill coefficient for the autoactivation.

For the simulation we used the parameter values from Table B.2. The value for the Hill coefficient for the self-activation was set to 2. The results of the simulation of the models with both the considered mechanisms for the dox-binding with added autoactivation terms are presented in Figs 3.9 and 3.10.

For the model with successive dox-binding we still have damping oscillations with the same 18 h period but with a three-fold increase in the amplitude of the first peak due to an increased amount of the rtTA produced by self-activation (Fig 3.9).

If we substitute the autoactivation term into the model with the simultaneous dox binding we immediately obtain the sustainable oscillations with a period of 8 hours and twice increased amplitude (Fig 3.10).

Bifurcation analysis revealed that the change in the model structure results in the expansion of the range for some parameter values that correspond to HB.

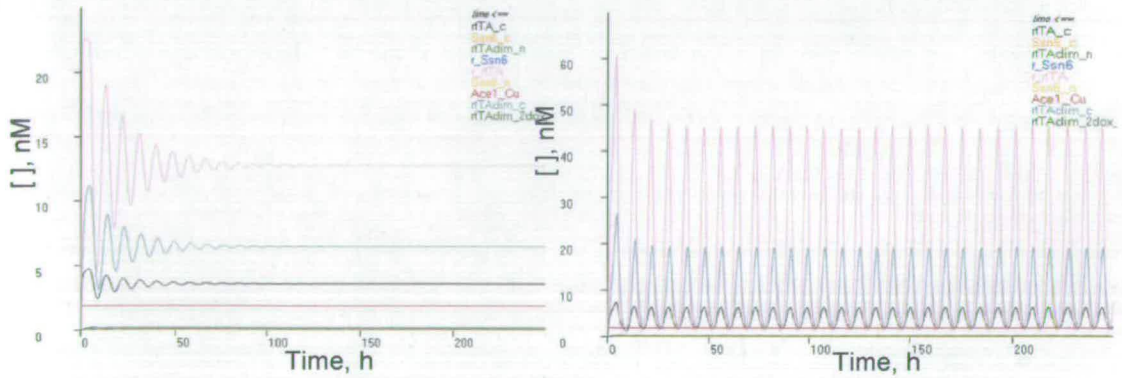


Figure 3.10 The results of the model simulation for the rtTA /Ssn6 synthetic circuit with the simultaneous mechanism of dox binding.
A. Time series for the initial model state.
B. Time series for the changed model structure. The system performs the sustained oscillations with the same 8h period and a two-fold amplitude increase

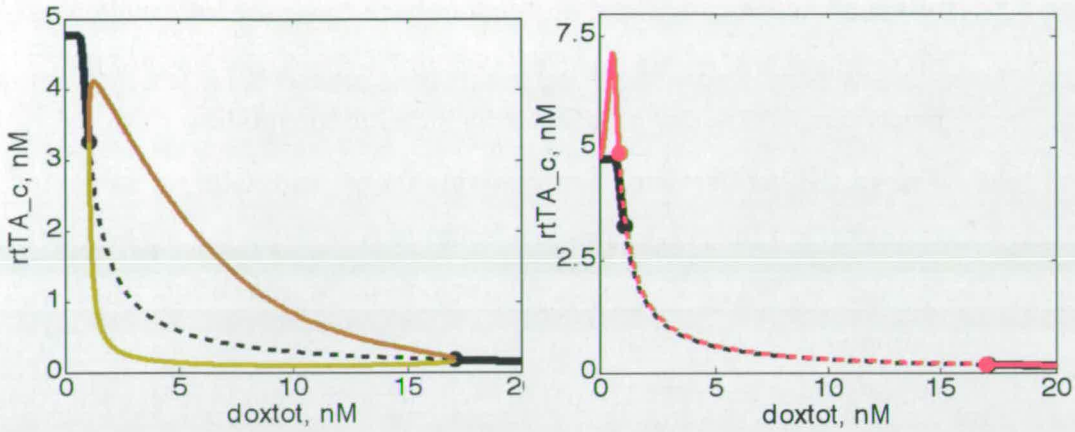


Figure 3.11 One-parameter continuation for the total concentration of the doxycycline (doxtot).

A. The diagram corresponds to the initial state of the model. The limit cycle is shown for rttA_c state.

B. The diagram corresponds to the model with a self-activation term.

For instance, the one–parameter continuation for the dissociation constant of the rtTa–dox complex demonstrates the greater allowed value for the constant in the case of the model with the autoactivation term (Fig 3.7, left curve).

Although there is no significant increase in the allowed range for total dox concentration (doxtot), which is clear from Figure 3.11, the one–parameter bifurcation diagram for the model with self activation (Fig 3.11, B) demonstrates the increase in the corresponding concentration of the rtTA_c state. Thus, the addition of self activation into the circuit would significantly increase the probability of detection of the oscillations, for it inevitably increases the concentration of the components and activator in particular. Also it generally allows wider ranges for the important parameters.

3.2.4. Introducing the external input

For a further introduction of the light input (Fig 3.4) we need to incorporate a new parameter in the model that will correspond to fluctuation of the transcription rate of one of the components caused by an external signal (light input). By varying the value of this parameter we will simulate the changing transcription rate depending on the strength of the external stimulus. What

kind of behaviour should we expect in this case and which parameters have to be tuned in order to bring the system back to sustained oscillations? To answer these questions we again used Oscill8 to check the possibility of the desired behaviour with a new parameter included in the model.

The initial model had a period of about 10h with a value for the basal transcription for rtTA (baseTR) equal to 0 (nM/h). Given the initial parameter settings in Table B.2 we generated a one-parameter bifurcation diagram on the parameter baseTR, changing the value for baseTR from 0 to 100 (nM/h). We found a HB at baseTR=15 (nM/h), which means that system oscillates within the domain [0, 15] nM/h for baseTR (Fig 3.12).

Thus, as a preliminary result, we have shown that the circuit can oscillate in the presence of some external force increasing the base transcription to 15nM/h level.

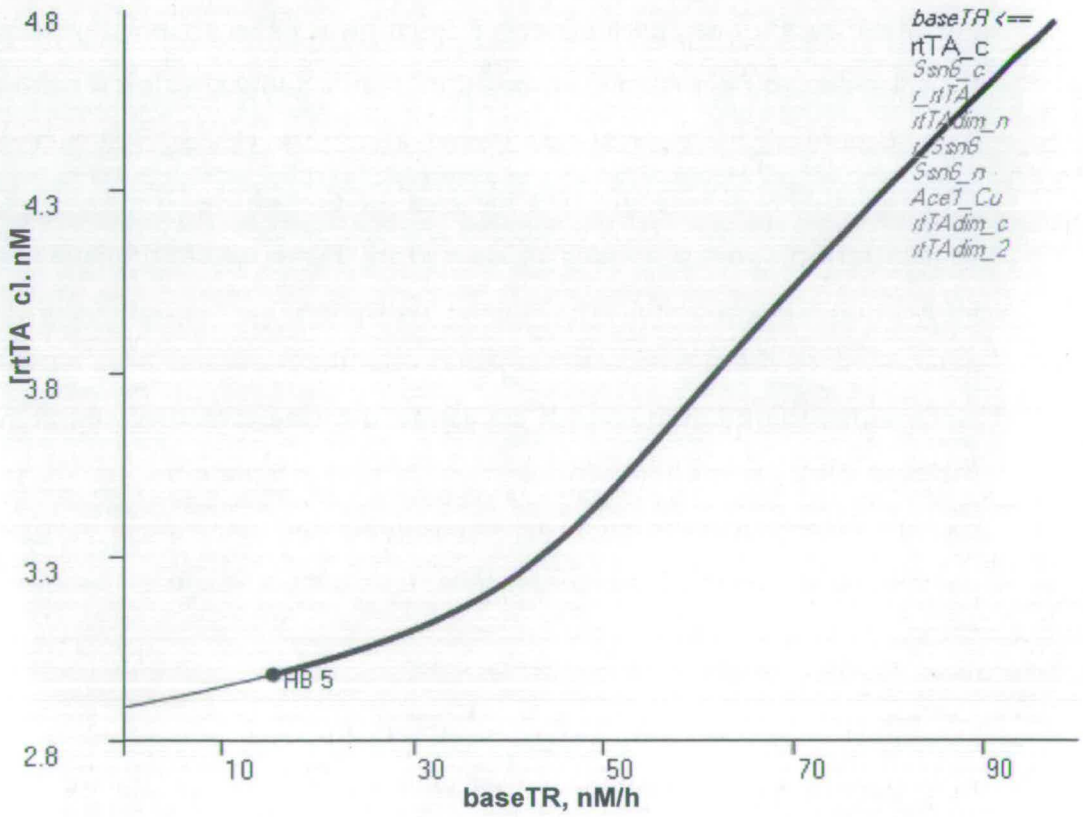


Figure 3.12 The one-parameter continuation for the basic transcription rate BaseTR .
The system performs the oscillator behaviour in the range for the baseTR [0, 15]

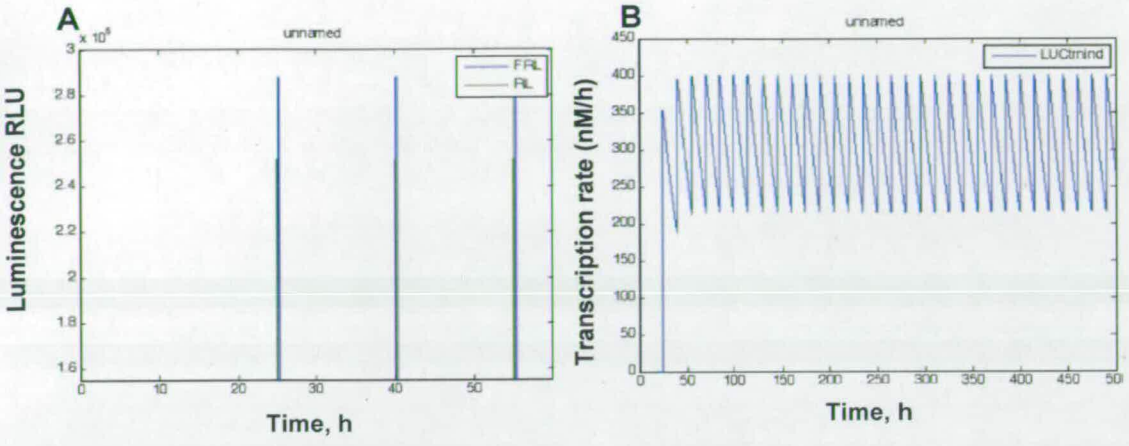


Figure 3.13 The results of the simulation of the effect of the R/FR treatment on the transcription rate.

A. Light protocol: 10min Ris followed by 15min FR light treatment.
B. The figure illustrates that the transcription rate in the light-induced subsystem is oscillating around a baseline of 300 nM/h.

However, the current model for the PHYA-based light input gives a much higher increase of transcription rate (Fig 2.7, B, C, and D). By the simulation of the light-switch model we revealed that the successive treatment by 10 min red and far red pulses with a 15 min interval between them significantly increases the baseline for oscillations (Fig 3.13). Therefore, I sought to find other parameter values that would allow the system to continue oscillating with the higher level of baseTR that would result from adding the light input module.

We proceeded with the two-parameter continuations of the HB with respect to other parameters in the system, searching for those parameter combinations that increase the value of baseTR at the HB. Two-parameter bifurcation diagrams were created for all parameters starting from the most important parameters for non-linearity: mainly Hill coefficients and degradation rates. The typical example of a two- parameter bifurcation diagram for basic transcription for rtTA (baseTR) and total concentration of the doxycycline (doxtot) is presented in Fig 3.14. Here, the total Dox concentration and `baseTR` are the competing parameters. The oscillations exist within the area which gives low `doxtot` with a relatively big `baseTR`, or a relatively big `doxtot` with low TRbase.

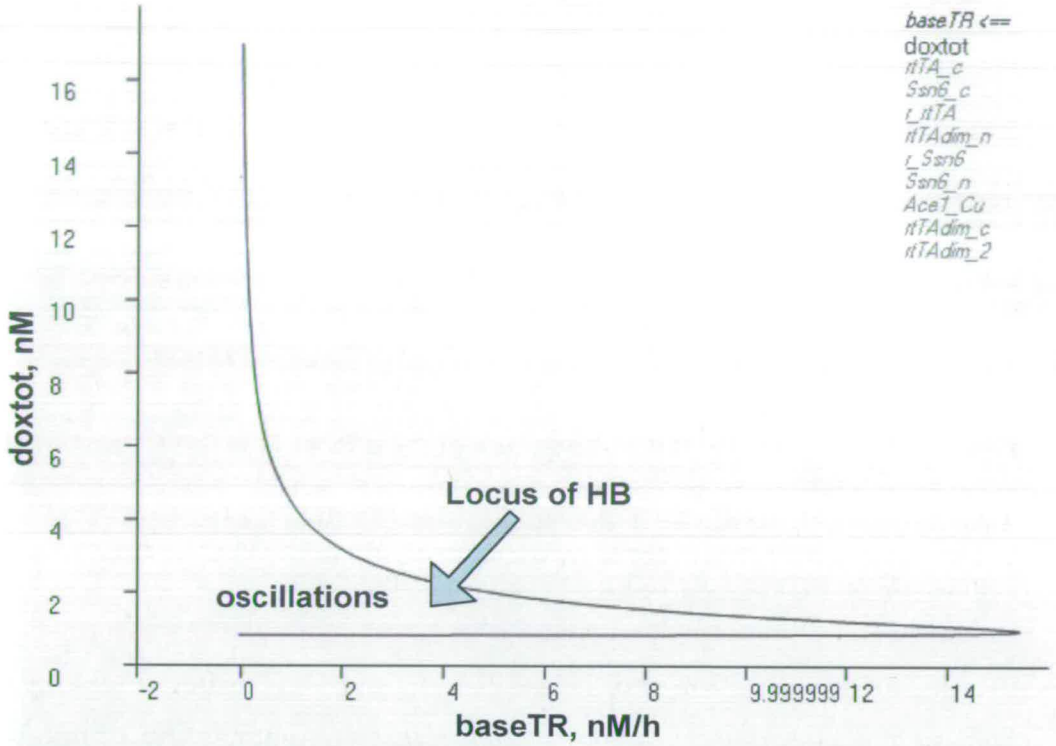


Figure 3.14 Two-parameter bifurcation diagram for the baseTR and doxtot parameters.

The solid line bounds the oscillatory region corresponding to the locus of the Hopf bifurcation.

As we are looking for the possibilities to increase the baseTR retaining the oscillating ability of the network, we proceed with two-parameter continuations for all the parameters with baseTR as a main parameter. We are searching for parameters that could be tuned in such a way as to get the substantial increase for baseTR. Modifying the values for n_rtTA (Fig 3.15), m_r_rtTA, Ki and Kdox_cplx we finally managed to enlarge the range for baseTR (Table B.3).

It was then necessary to adjust the parameter values changed during the manual tuning back to biologically sensible ranges. The procedure was the same as described above and took about 100 iterations and about 1000 two-parameter continuations generated to get the final oscillatory system (see the parameter values in Table B.3). The model obtained in the above study has an

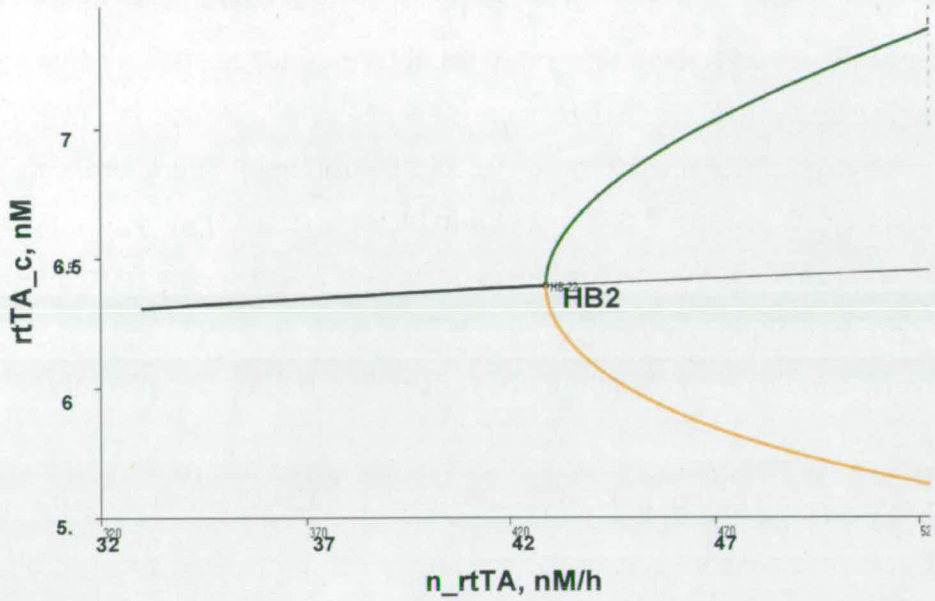


Figure 3.15 The bifurcation diagram for parameter n_rtTA.
Again, as at the Figure 3.7 HB2 is a supercritical Hopf Bifurcation. n_rtTA corresponds to the maximum rate of transcription for rtTA. It shows that when the parameter attains the value of about 430, it meets the HB point and the

internal period of oscillations T=12.5 (h). It maintains the sustained oscillations within the range of baseTR [100, 301] nM/h.

3.2.5. Investigation of model oscillatory potential with an inverse bifurcation analysis

In Systems Biology modeling, detailed kinetic knowledge on the parameter values is rarely available. Even the knowledge of the network topology is a matter of making assumptions, as in our case. However, for a given gene regulatory network, the principal question arises at the very first modelling step: can the set of genes with their hypothetical interactions produce bistable or oscillatory behaviour for some particular choice of parameter values? And, if yes, which of the minimal choice of parameters could be tuned to attain the desired dynamical performance?

Questions of this nature could be qualified as an inverse problem when, in the absence of prior knowledge on parameter influence, the system could be brought to the relevant parameter regime and the initial guess on the importance of different parameters could be obtained. The inverse eigenvalues analysis (IEA) proposed by James Lu in [154], and [155] provides a very useful facility for solving inverse problems of this kind and seems particularly useful for Synthetic Biology where someone may wish to examine a set of possible network topologies and eliminate those that fail to behave in the desired manner.

The strategy suggested in [154], and later in [155], includes two main steps:

1) Use inverse eigenvalue analysis (IEA) to see if a model can exhibit desired qualitative behaviours, e.g., oscillations or bistability. IEA allows probing and characterising of the possible combinations of parameters, which lead to changes in qualitative dynamics. Figuratively speaking, IEA represents a very convenient way to form the shortcut to the nearest Hopf bifurcation in a high dimensional parameter space.

2) Once a bifurcation has been found, then one can engineer the model for desired characteristics, e.g., the value of parameter leading to oscillations or robustness of the period of oscillation. Generally, inverse bifurcation helps to identify and design mechanisms that condition to certain bifurcational phenotypes.

In this work we used only the IEA for analysing the possibility of sustained oscillatory behaviour for our topology designs.

The strategy of IEA is based on computing the Jacobian matrix corresponding to the Hopf bifurcation (see 3.2.2. for the HB conditions). In the proposed algorithm the matrix inverse eigenvalues problem has to be solved, where the Jacobian matrix of interest is calculated, and where the corresponding upper and lower bounds and linear constraints for all the entries are obtained from the reaction network structure and kinetic laws prescribed for each reaction. The second step involves a nonlinear constrained optimization problem, where those parameter values and the state solution have to be found that match the

a-priori solution with a minimal divergence. Simply, we have to minimize the distance between the current point in parameter space and the closest Hopf bifurcation manifold, so as to bring the system to the boundary of the stable regime[154, 155]. The process includes multiple optimization iterations, each of which requires a number of one–parameter continuations[154, 155]. Further details of the algorithm can be found in the James Lu paper [155] which in particular includes the case illustrated with the version of the model for simultaneous dox-binding mechanism presented here .

We were using the Inverse Eigenvalues Toolbox in Mathematica to implement the algorithm introduced above. The Toolbox was kindly placed at our disposal by James Lu (personal communication).

We tested two of our principle topologies, Gal4-based I and Tet-based, to identify parameter sets leading to the sustained oscillations. As the biologically meaningful range of parameters has been obtained from the literature, it is also necessary to determine whether parameters lying in the allowed range can satisfy the problem specification.

For the Tet-based system we use the initial parameter values with the allowed range presented in Table B.2.

The results of IEA applied to the models with simultaneous and sequential dox binding mechanisms are presented in Figure 3.16

Table 3.1 represents the resulting parameters obtained with IEA. Interestingly, in all the considered cases the two parameters that most strongly contribute to oscillations are the dissociation constant for the dox-rtTA (Kdox_diss) complex and Michaelis constant for the transcription of inhibitor Ssn6 (g_Ssn6). Recall that during the forward bifurcation analysis we found the oscillatory behaviour for the model with simultaneous dox binding mechanism by tuning the same constant Kdox_diss.

Therefore, the IEA gives similar results in comparison with forward bifurcation analysis in the situation where we have random initial guess for parameter values. Moreover the IEA significantly facilitates the attainment of the nearest HB in highly dimensional parameter space without trying multiple insufficient

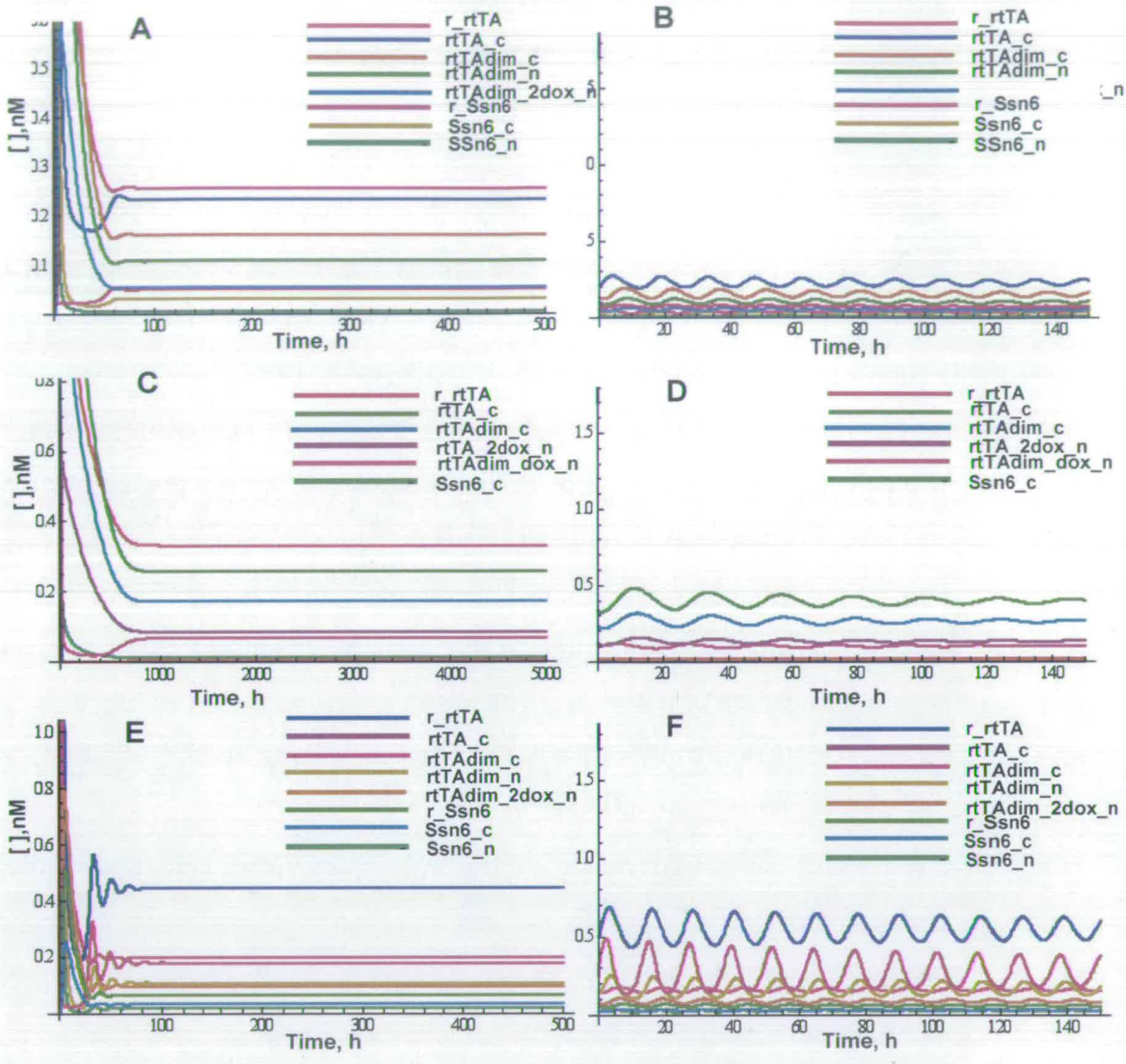


Figure 3.16 The dynamic behaviour of the Tet-based model system before (left) and after (right) the application of the IEA method.

A. The model describing simultaneous dox binding with initial parameter values. B The model describing simultaneous dox binding after the IEA with the parameter values modified as presented in Table 3. 1. Resulting period of oscillations is about 18h C. The model for sequential dox binding before the IEA upplication. D. The model with sequential dox binding after IEA with the parameter values tuned as presented in Table 3.1 Resulting period of oscillations is about 20h. E. The model with simultaneous dox binding with an additional parameter for the basic rtTA transcription (baseTR) before the IEA. F. Model with base TR after IEA with resulting period of about 12 h.

Table 3.1 Parameter values obtained with IEA

		simultaneous dox	sequential dox	simutaneous dox with baseTR
param names	initial value	IEA	IEA	IEA
K_dox_diss	0.07 (h^{-1})	0.19 (h^{-1})	0.66 (h^{-1})	0.29 (h^{-1})
g_Ssn6	1 (nM)	0.6 (nM)	0.02 (nM)	0.28 (nM)
m_r_rtTA	4.33 (h^{-1})	4.29 (h^{-1})	4.28 (h^{-1})	4.29 (h^{-1})
a_Ssn6	2	2	2	2.02
m_r_Ssn6	3.85 (h^{-1})	3.82 (h^{-1})	3.81 (h^{-1})	3.81 (h^{-1})
p_rtTA	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})
a_i_Ssn6	4	4.02	3.97	4
m_Ssn6_c	1 (h^{-1})	0.99 (h^{-1})	1(h^{-1})	1 (h^{-1})
m_rtTAdim_n	5 1 (h^{-1})	4.98 1 (h^{-1})	4.85 1 (h^{-1})	4.98 1 (h^{-1})
n_Ssn6	10 (nM h^{-1})	10.02 (nM h^{-1})	10 (nM h^{-1})	10.03 (nM h^{-1})
K_dox_cplx	1.19 (nM^{-1} h^{-1})	1.18 (nM^{-1} h^{-2})	1.18 (nM^{-1} h^{-1})	1.18 (nM^{-1} h^{-2})
Kdim_diss	1 (h^{-1})	0.99 (h^{-1})	0.99 (h^{-1})	1 (h^{-1})
p_Ssn6	1 (h^{-1})	1 (h^{-1})	0.99 (h^{-1})	1 (h^{-1})
K4	1 (h^{-1})	0.99 (h^{-1})	1 (h^{-1})	1 (h^{-1})
Cutot	8 (nM)	7.99 (nM)	7.9 (nM)	8 (nM)
doxtot	2 (nM)	2 (nM)	1.99 (nM)	2 (nM)
Ki	0.01 (nM)	0.009 (nM)	0.017 (nM)	0.01 (nM)
K1	1 (h^{-1})	1 (h^{-1})	0.99 (h^{-1})	1 (h^{-1})
K3	1 (h^{-1})	0.99 (h^{-1})	0.99 (h^{-1})	1 (h^{-1})
Kdim_cplx	1 ($nM^{-1}h^{-1}$)	0.99 ($nM^{-1}h^{-1}$)	1 ($nM^{-1}h^{-1}$)	1 ($nM^{-1}h^{-1}$)
K2	1 (h^{-1})	0.99 (h^{-1})	0.99 (h^{-1})	2 (h^{-1})
a_rtTa	6	6	5.99	6
g_rtTA	1 (nM)	1 (nM)	0.99 (nM)	1 (nM)
Ace1tot	40.36 (nM)	40.359 (nM)	40.35 (nM)	40.36 (nM)
n_rtTA	100 (nM h^{-1})	100 (nM h^{-1})	100 (nM h^{-1})	100 (nM h^{-1})
K_Cu_cplx	1 ($nM^{-4}h^{-1}$)	1 ($nM^{-4}h^{-1}$)	1 ($nM^{-4}h^{-1}$)	1 ($nM^{-4}h^{-1}$)
K_Cu_diss	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})
m_Ssn6_n	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})	1 (h^{-1})
baseTR	0.1 (nM h^{-1})	-	-	0.04 (nM h^{-1})

one-parameter continuations. It may happen additionally that none of the axes in Oscill8 crosses the bifurcation manifold (and this becomes more probable

with the growing parameter number), so the one-parameter search would be completely ineffective in this case.

Unfortunately, as the IEA is looking for the nearest HB, the success of the search depends heavily on the initial guess, namely, the proximity of the initial point to the destination. In our case, we failed to localise with IEA the HB for the value for the baseTR bigger than 1 nM/h. This means that the given initial guess is quite distant from the desired point. Thus, additional iterations in Oscill8 for revealing the relationships between the parameters would be required for targeted parameter adjustments.

In the light of the above discussion, the proposed Tet- based system appears accessible for further implementation and study as it gives multiple possibilities for oscillatory behaviour depending on parameter values. The availability of experimental data will give the opportunity to choose the appropriate model mechanism from the several suggested. However, in the meantime we can perform *in silico* experiments for circadian clock entrainment using the best of the basic models: the model for simultaneous dox binding mechanism.

3.2.6. Simulation of entrainment of core oscillator

Every circadian clock system has to detect external signals and light to be able to adjust itself to the daily changing environment. The input pathway allows the entrainment of the core oscillator, one of the key characteristics of circadian clock. This generally implies that the external signal can reset the clock by inducing transcription or promoting the degradation of the particular clock component. From the mathematical point of view the resetting corresponds to the recurrence of the limit cycle after the application of perturbations. Depending on their strength and time of application, the perturbations may cause the so-called phase shifts when the state variable moves back to the limit cycle with a different phase compared with non-perturbed conditions [156].

In subsection 3.2.4 we have introduced the parameter baseTR to our model for mimicking the inducing effect of the external forcing in the form of light input on the transcription rate of the activator. With our model core oscillator we can investigate how these particular oscillatory systems will respond to the fluctuations in transcription rate caused by the external forcing. We also can try to vary different characteristics of this external forcing, starting from a simple On-Off step function and finishing with our own modeled and implemented PHYA-based light input.

a) *The simulation of entrainment with the square shape external forcing.*

We start from application of the direct influence of the external forcing that has a form of a step function and could be realized by a corresponding perturbation of the base TR parameter. For simplicity it will be called the *square shape external forcing*. In the first *in silico* experiment we applied the external forcing of different periods from 5 to 15h. The phase portraits of the entrained system are presented in Figure 3.17. The figure shows the dependency of one state of the system (rtTA) on another (Ssn6). In an ideal oscillator such dependency would have the shape of a thin closed curve, as a combination of the state variables will repeat exact values in consecutive periods. Any deviation from one thin line indicates disruptions of the oscillator, thus, images like those shown for period 9 would indicate poor entrainment.

As it seen from the figure, the system with an internal period of 12.5 h could be easily entrained to an external period of 11 to 16h. Smaller (6h) and large (24h) periods could be reached after some time, but the system will oscillate with erratic amplitude.

b) *Free run- external 15h- free run experiment*

After we simulated the effect of the external forcing on the dynamics of our core oscillator we decided to examine how the system would behave if the external force is applied during some transient period of time. Basically, the question under the consideration is whether the system would prefer to retain the period obtained by the influence of the external forcing or if it would quickly return to the original period.

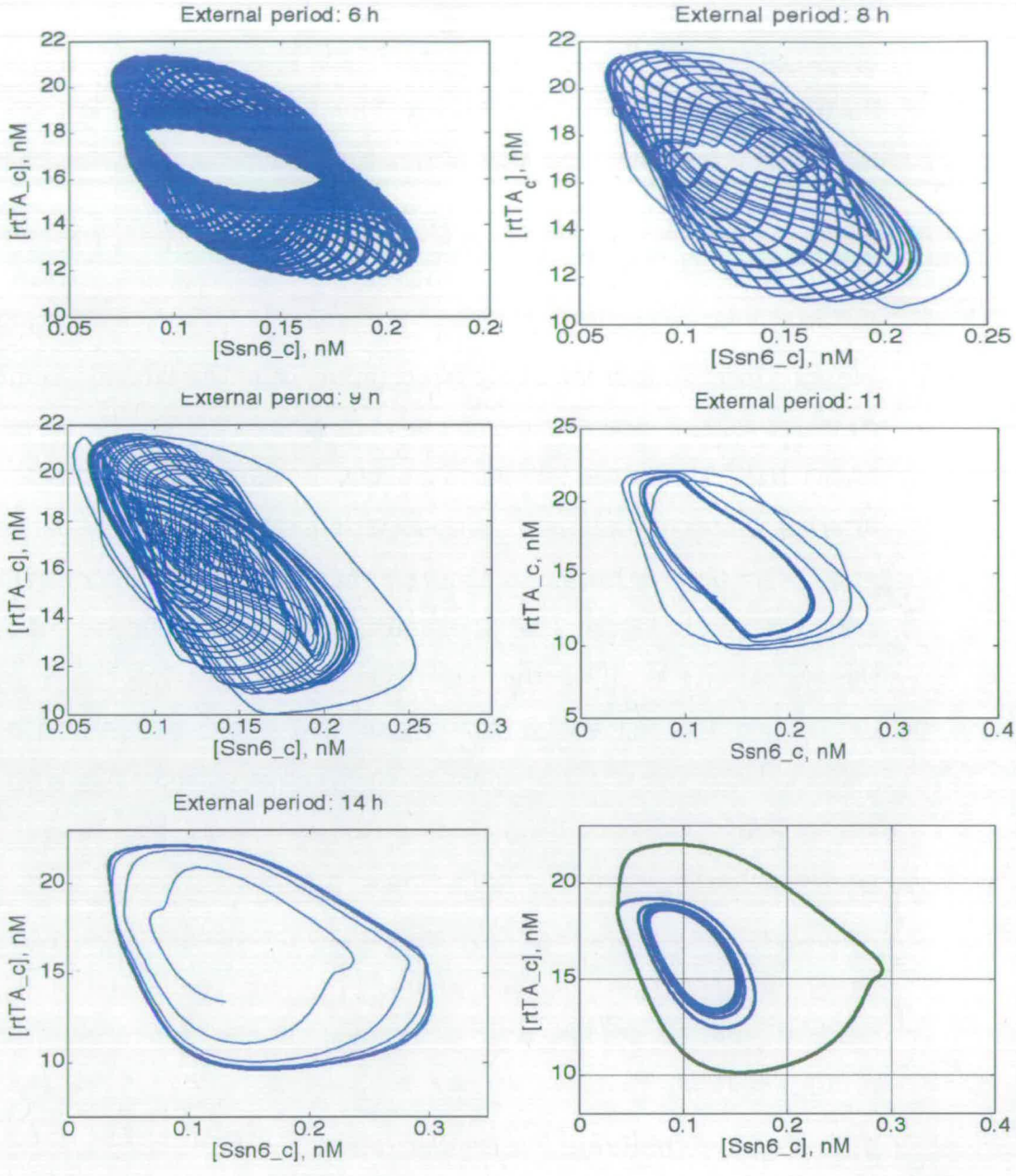


Figure 3.17 The results of simulation of the entrainment of the core oscillator with the square shape external forcing of period from 5 to 15h.

The lower right figure represents the transition of the initial period of 12.5 hours to the one resulting from the external forcing of 15h period.

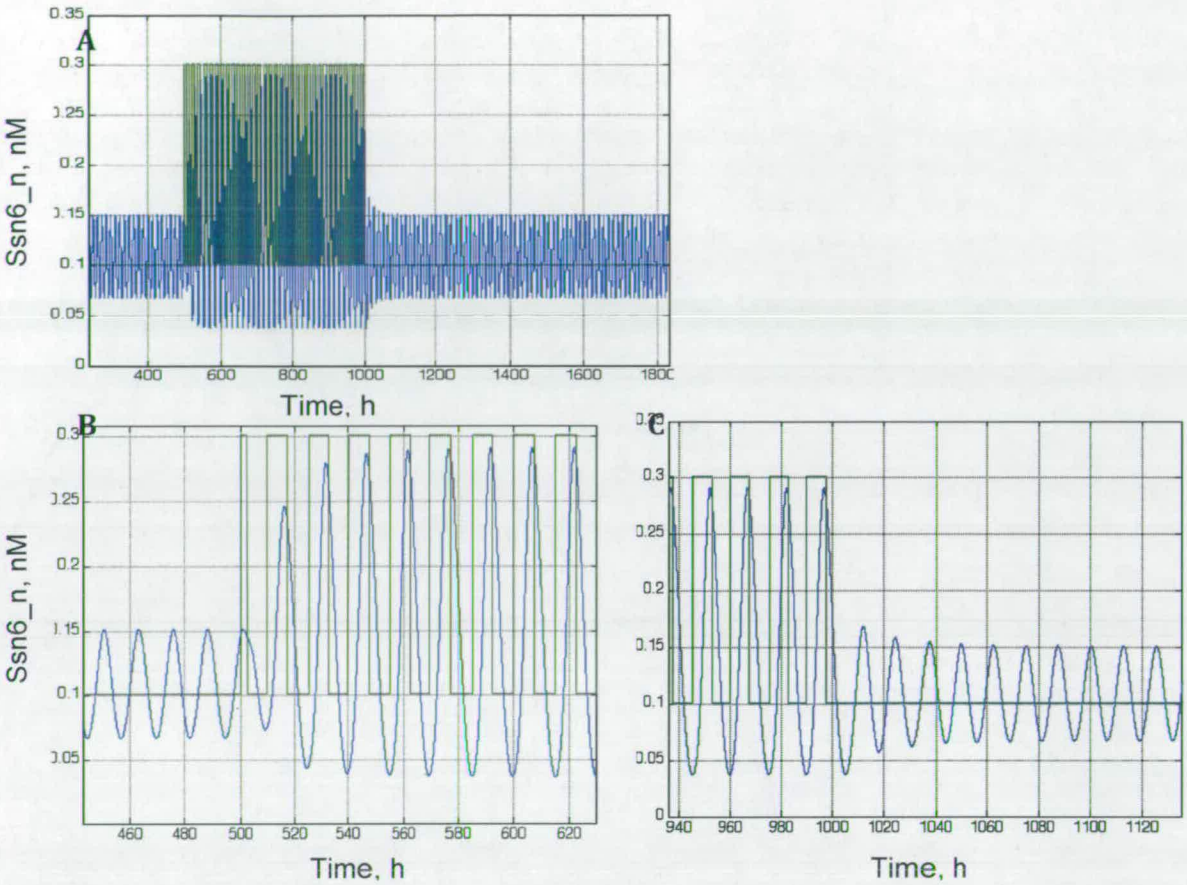


Figure 3.18 The results of simulation of the transient effect of the external forcing of the period of 15h.

A. The square shape external forcing was applied at time 500h and removed at time 1000h B. 6 cycles are needed to bring the system oscillating with a period of 12.5h to the period of 15h. C. After removal of the external forcing the system immediately returns to the original period.

To perform this experiment we applied the square-shape 15h period external forcing that has been shown to be sufficient for entrainment in the previous simulation. The signal was applied to the oscillator with the internal period of 12.5h at time 500h. After 6 periods the system had adopted a 15h period and began to oscillate with the external forcing period. Release from the external forcing happened at time 1000h with an immediate return to the initial period of 12.5h (Fig. 3.18)

c) *Simulation of phase shifting with the square shape impulses*

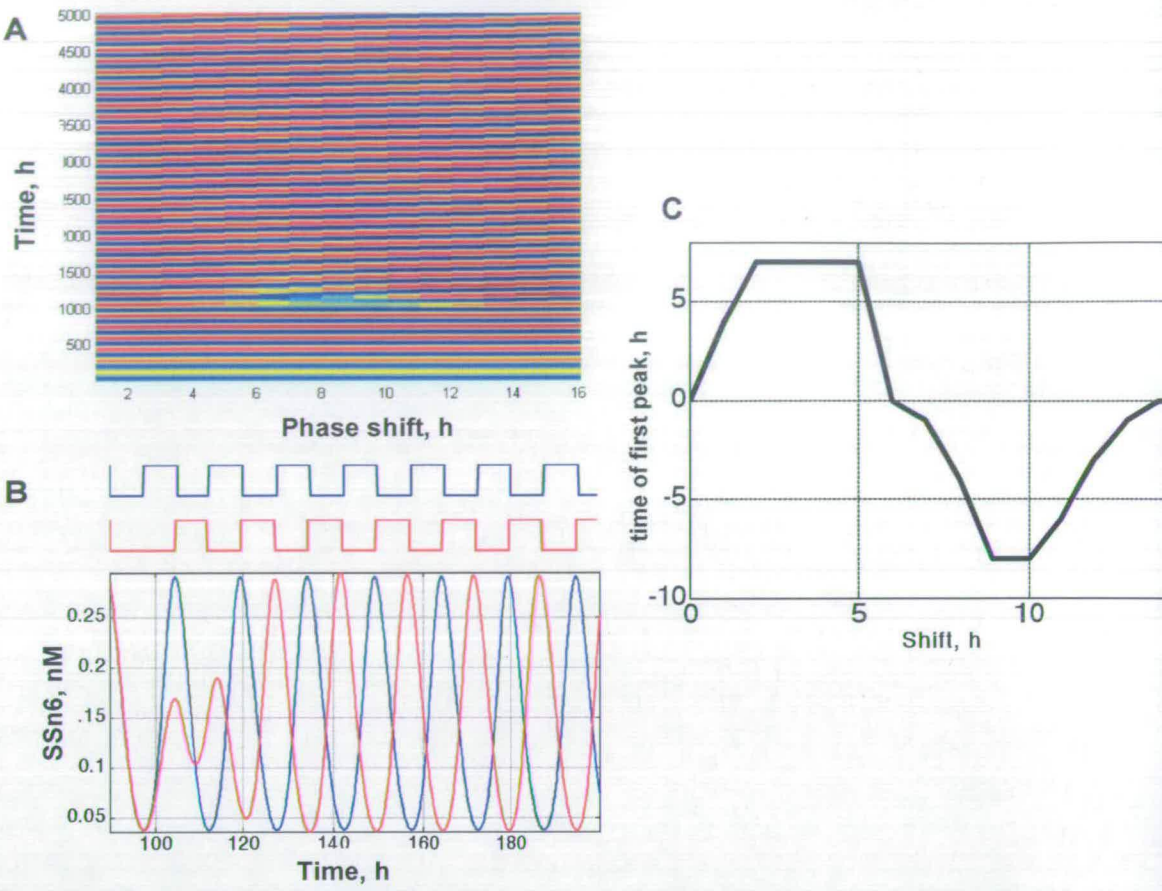


Figure 3.19 The results of simulation of the phase shifting resulting from a delay applied at different times of the external cycle

A.2D map of phase response to phase-shift applied at different times of the15h cycle is presented. Pseudocolors represent concentration of Ssn6, ranging from 0 nM (light blue) to 0.3 nM (red).

B. Phase delay applied at the midpoint of the 15h cycle has a pronounced effect on the next two cycles (blue colour corresponds to unperturbed system, red corresponds to perturbed system). The protocol of treatment is presented above the picture in respective colours.

C. A kind of PRC could be created with the data above: we plot the time difference between the expected and actual first peak after the phase shift applied at different times of the 15h external cycle.

We simulated the phase shifting resulting from the application of the external forcing with the same characteristics (15h period) at different times of the internal cycle of the system. For this experiment the system was initially entrained to an external 15 h rhythm with a 7.5/7.5 –"light/dark "cycle. Phase shift in the form of delay (jet lag) has been applied at different times of the 15h cycle. The Figure 3.19 shows that the delay applied close to middle of the T, 7h,

has a more pronounced effect, while the phase shifts at 2 and 12 h do not greatly affect the behaviour of the system.

d) *The simulation of entrainment with the triangular shape external forcing.*

Light-induced transcription has a *triangular* rather *than* square shape as one can see at Figure 3.13. To investigate influence of the signal of a more realistic shape on the transcription rate and corresponding oscillatory performance, we have applied the simulated external forcing of a triangular shape to our oscillator. This kind of signal differed in its resulting effect on period compared with the square shape, but still the system could be stably entrained to a period from 10 to about 17 h (Fig 3.20, A). Interestingly, the oscillator gains the second and third harmonics when being entrained to 15h of the triangular shape external forcing (Fig 3.20, B). This effect strongly depends on the strength (in terms of amplitude) of the applied forcing and appears when the amplitude of the signal exceeds the value of 50 nM. The erratic shape of the phase portrait is also determined by these emergent harmonics.

In real biological systems the effect of light input decreases over a time period much longer than the period of the oscillator. Therefore, any shift in a sequence of light pulses will cause not only changes in the phase of the oscillator, but will have an impact on the amplitude over a substantial amount of time (Fig 3.20, C). The magnitude of this change of the amplitude will strongly depend on the time between the consecutive pulses at the moment of phase shift. In our modelled system, the change in the amplitude of the oscillator's response is much higher within a 2 h shift than within a 15h shift

We investigated the range of effective input signal strengths in terms of the amplitude of the input signal (Fig 3.20, D). The figure illustrates the combined effect of amplitude and period of external forcing on the period obtained. Generally, the system could be well entrained to the period from 11 to 15 h. However, a relatively high (from 70 to 100 nM) effective amplitude of external signal is needed for entrainment of the system to the period of 14-15h, while in the case of periods close to the internal cycle (12.5h) a much wider range of amplitudes works sufficiently.

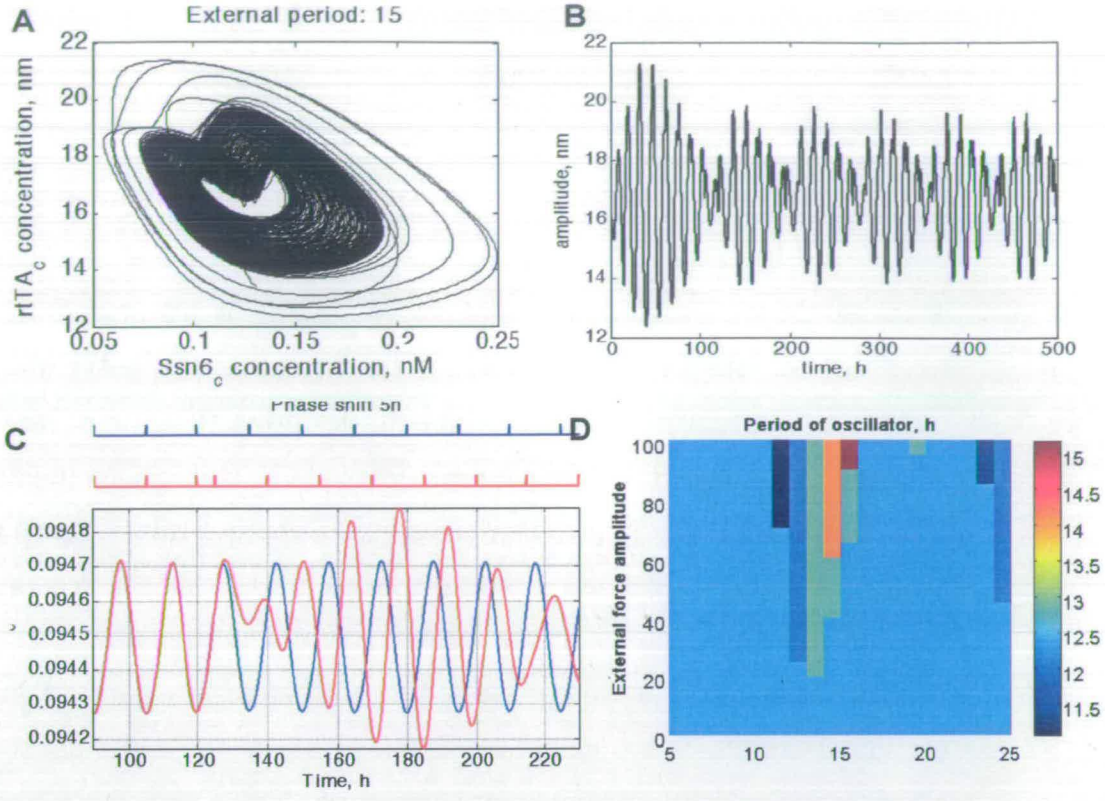


Figure 3.20 The results of simulation of entrainment of the core oscillator with an external forcing of with a more realistic (triangular) shape.

A. The phase portrait for the 12.5h internal cycle entrained to the 15h external cycle.

B. The oscillator gains the second and third harmonics when being entrained to 15h of the triangular shape external forcing with amplitude of 80 nM.

C. The phase-shifting experiment: the oscillator is entrained to 15h. After the time point 120h the next signal comes after 5 h instead of 15h

D. The 2D diagram for the resulting period obtained by the oscillator, which depends on the period of the applied external forcing (axis X) and its amplitude (axis Y)

The examples for different periods from the diagram (Fig 3.20, D) are presented in Figure 3.21 with the amplitude of the applied external forcing drawn along the X-axis and the resulting period presented along the Y-axis (period was estimated as a period of Fourier component with maximum coefficient). As can be seen, not all the periods of external forcing equally allow the entrainment of the oscillator in the proposed range of strength. For instance, for periods between 11 and 15 hours there is a applied force amplitude threshold above which the system oscillates with the period close to the period of external forcing, while the external forcing with a period of 19h, for example, just

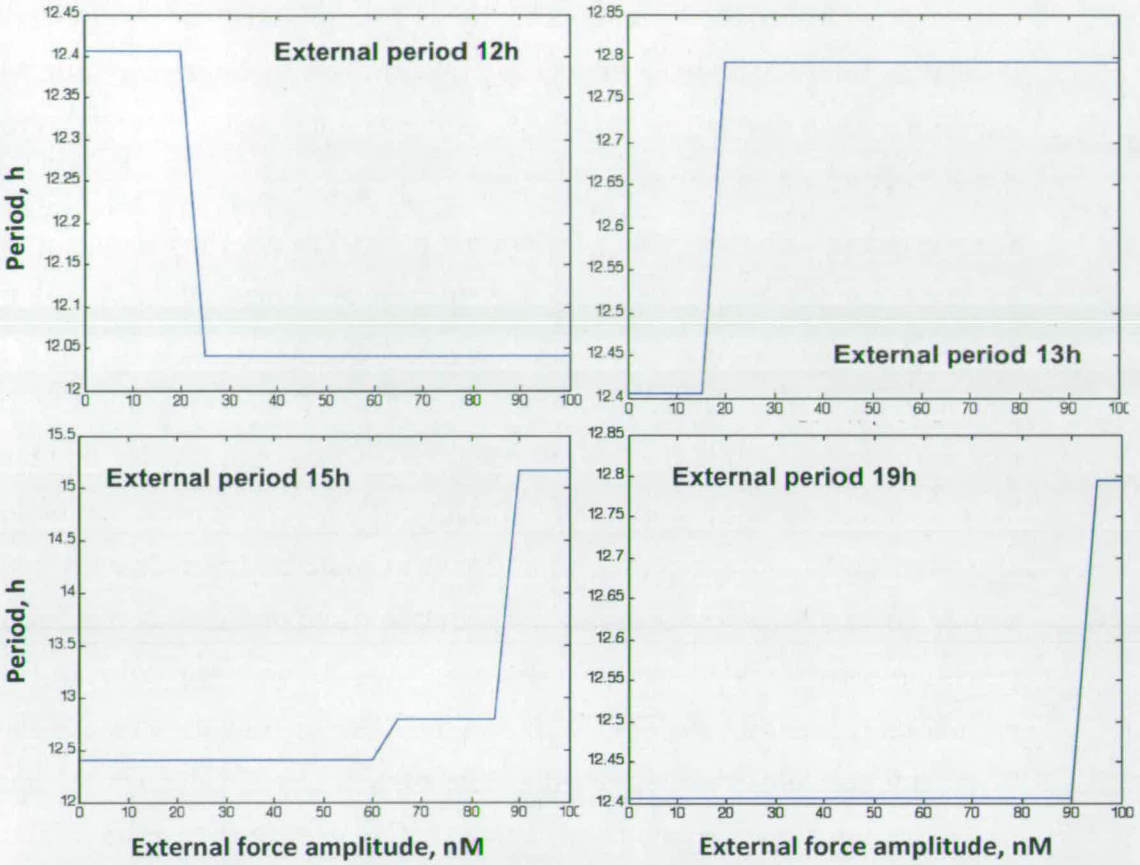


Figure 3.21 The effect of different characteristics of the applied external forcing on the resulting oscillatory period.

slightly prolongs the original period. Forcing cycles in a range of 22 – 25 h decrease rather than increase the oscillation period to the half of the period of the external forcing, owing to the phenomenon of 2:1 entrainment (Fig. 3.20, D).

3.3. Discussion

We examined the two suggested designs for the synthetic core oscillator by preliminary mathematical modelling. The first one is based on the well-known yeast GAL regulon. Some kinetic data for this system, as well as a few mathematical models for Gal4-Gal80 relationships are available in the literature. Those models were developed to describe in mass action kinetics form the cascade of events from adding of the galactose to the media to the point of free GAL4 release. They include the GAL4/GAL80 and GAL3/GAL80 interactions and formally describe the distribution of GAL80 within the cell, so some parameter data could be adopted from these models, although they do not consider any

transcription/translation events [132, 135, 136]. However, the incorporation of these data into our phenomenological mathematical model resulted in switch-like rather than oscillatory behaviour, which is consistent with the literature, but doesn't fit our purposes.

We proceeded with the second, Tet-operon based, design. The bacterial origin of its main components gives a strong advantage compared to a GAL-based system, as it allows the minimizing of unavoidable cross-talk with a host yeast cell. We proposed a mechanistic model, describing all the biochemical reactions that may happen within the system. Two possible mechanisms for dox-binding have been considered separately. Both of them correspond to boundary situations and a more complicated mechanism could be approximated from the simplified one if corresponding values for the rates of forward and backward reactions satisfy the inequality calculated in 3.2.1 and Appendix C. For the parameter values that were derived from the literature and used for simulations it was not possible to decide whether this inequality was valid or not, thus, the two cases have been treated separately. The *in vivo* implementation and real experimental data that would be obtained for the system may change the presumed parameter values, and for the real system the model will be chosen that matches the real situation.

We used two main strategies for the exploration of the oscillatory potential of the proposed topologies: the forward and inverse bifurcation search. Both were successful and gave nearly the same results. In fact the initial parameter guess for the given topology appeared to be sufficiently close to the Hopf bifurcation manifold (HB). Nevertheless, the use of an Inverse Eigenvalue Analysis (IEA) significantly facilitates the search for oscillatory behaviour as it allows the automatic attainment of the HB without strong prior knowledge about parameter values. The best strategy would be the combination of both the forward and inverse approaches: starting from the localization of the nearest HB with the IEA and then proceeding with a continuation for the parameter of interest in some tool like Oscill8, then taking the value close to the desired one and returning to the IEA to find the oscillations. For synthetic biology purposes

such a strategy is particularly advantageous as it allows the exclusion of ineffective topologies at the preliminary steps and a focus on the network parameters that mostly contribute to the desired behavior.

Therefore, we managed to localize the oscillatory regions in the parameter space for all models, staying at the same time within biologically reasonable ranges for all parameters. At this stage the system is ready for *in vivo* implementation and further modeling or experimental iterations

We also investigated the system's capability to respond to external signals, as occurs in natural circadian clocks. For this we introduced an additional parameter that corresponds to the basic transcription rate for the activating component. Modulating this parameter in a different manner we showed that our Tet-based system could be entrained by an external signal of varying shape, amplitude and period. Through *in-silico* experiments we examined the possibility for the modelled core oscillator to interact with our built and modelled light input system.

At the next step, with the practical construction of the core oscillator in yeast, this artificial system will be able to facilitate clock–related investigations.

Chapter 4. Modelling of clock output

We have chosen *Ostreococcus tauri* as model organism for modelling of possible clock regulation of the starch metabolic pathway. We expect, based on literature data for *Galdieria sulphuraria* and *Chlamydomonas reinhardtii,* that the post translational regulation and especially the redox regulation of the starch metabolic pathway is not so pronounced in *Ostreococcus* as in higher plants

We propose that if metabolic fluxes in *O.tauri* are driven mainly by transcriptional regulation they could be directly approximated from gene expression data sets under diurnal conditions. In fact, genetic processes are much slower than metabolic ones. We assume that all regulatory feedback influences will be equilibrated rapidly enough to treat the flux system as being in quasi-steady state on a time scale of three hours. Thus, the behaviour of the system of interest could be approximated as set of steady states with rapid transitions between them. These steady states will correspond to the genetic states available through the microarray data. In the absence of detailed dynamic information about regulation of the metabolic network by the circadian clock such an approximated model could be a useful tool for unveiling of genetic-metabolic interactions.

There are several ways to model metabolic systems, ranging from less complicated stoichiometry-based models to detailed kinetic models. Among the most popular is flux balance analysis (FBA), which is widely used for the prediction of flux distribution with linear programming methods. Based on the knowledge of reaction stoichiometry, various constraints and biomass composition, FBA has been applied in investigations of the full metabolic model for several organisms including *E.coli* [157, 158], *S.cerevisiae* [159] and *C. reinhardti* [160] These models are useful in predicting maximum growth yields and for rational strain design, as well as for viability of knockouts.

When applied to a particular pathway, FBA allows the prediction of pathway capability, and, by *in silico* deletions and overexpression analysis, interrogates the potential regulatory steps and, hence, candidate drug targets [161, 162].

The process of FBA modelling starts from metabolic reconstruction, which results in the creation of a stoichiometric matrix. The latter is used for further analysis with generally applicable methods such as extreme pathway and robustness analysis [4], and also with methods specific for the existing data type and the problem posed [161, 163]. Additional information available from transcriptome data can significantly improve the results obtained by FBA. A few successful approaches to incorporate the gene expression data to the flux model to infer the metabolic information have been done during the last decade. Among them the Boolean –like tri-value mapping of gene expression states for revealing of the tissue –specific enzyme activity reported by Shlomi et al [163], which allowed accounting for post-transcriptional regulation even in the absence of biologically inspired objective function. The alternative method called E-flux has been proposed recently [164]. It operates directly with reaction constraints setting the maximum flux capacity in accordance with the gene expression level. This approach was successfully used in drug target predictions for *M. tuberculosis.* Although it did not take in consideration the post-transcriptional regulation, it was not that important for analysis of steady-state reached by bacteria in response to environmental and genetic perturbation. Therefore, both approaches mentioned were used for inferring static metabolic information while the applications to dynamics of the metabolic state are still limited.

By accomplishing a full reconstruction of the *O.tauri* starch metabolic pathway, we created a detailed reaction model, which describes all the steps of starch biosynthesis and degradation known for the Choroplastida. We integrated gene expression (microarray) time series for *O.tauri* to the flux balance model of starch metabolism in order to test the principle transcriptional regulatory pattern. The method proposed allows the prediction of the flux distribution and temporal behaviour of key metabolites (starch, maltose and glucose 1-phosphate) during the 12/12h light/dark cycle. Single deletion and overexpression analysis reveals the main regulatory steps, responsible for the diurnal starch pattern and, thus, potential targets for circadian clock regulation.

4.1. *O.tauri* lacks redox regulation of starch metabolism

First of all, based on literature and sequence databases (Uniprot), we checked whether starch metabolism in *O. tauri* demonstrates the same potential complexity in posttranslational regulation as plants do. We were looking for the enzyme target sequences responsible for the redox-related modifications similar to those that have been described in literature for plant species.

As we expected, like *C. reinhardtii, O.tauri* seems to lack the potential for redox regulation at least at the level of starch metabolism that could be easily illustrated by the examples of the AGPase, GWD and β-amylase enzymes.

When oxidised and dimerized the small subunit of ADP-glucose phosphorylase (AGPase) is inactive, resulting in low affinity and decreased accessibility for allosteric regulation. In plants dimerization occurs through the disulfide bridge via Cys-82. The corresponding QTCL motif is well conserved in the N-terminal region of AGPB genes of almost all dicots and some monocots [91]. However, this motif is absent in *E.coli*, cyanobacteria and *Chlamidomonas reinhardtii* [100, 165]. Multiple alignment for the small AGPase subunit for the *Ostreococcus* and *Micromonas* species reveals that none of them has the QTCL motif needed for Trx regulation (Fig 4.1).

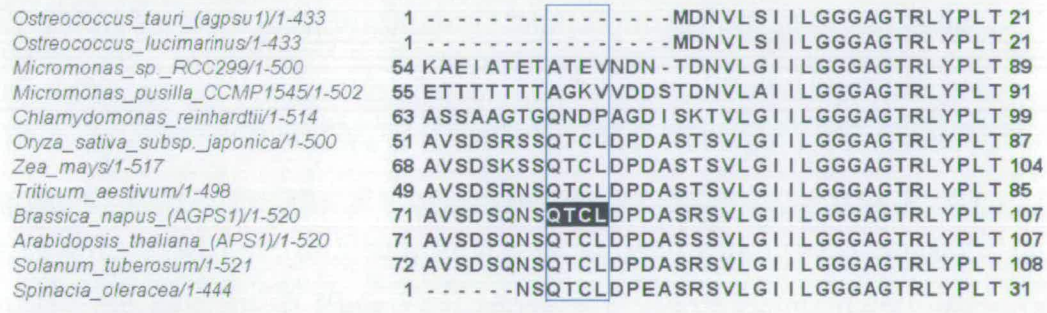


Figure 4.1 The multiple alignment for the small (beta) subunit of ADP – glucose phosphorylase

***Arabidopsis thaliana, Solanum tuberosum*, pea, wheat, maize, *Ostreococcus tauri* and *lucimarinus, Chlamydomonas reinhardtii* and *Micromonas pusilla* and *CCMP490*. The sequence QTCL characteristic for the higher plants is absent in unicellular algae.**

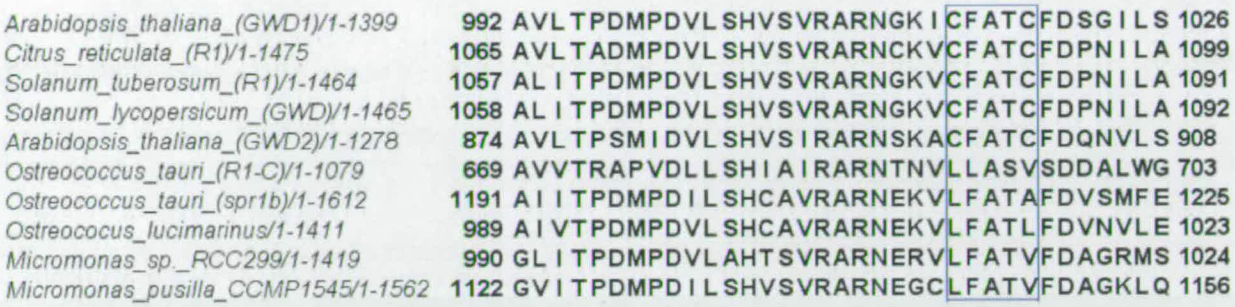


Figure 4.2 Multiple alignment for GWD1 (α-glucan, water dikinase) from *A. thaliana, S. tuberosum, C. reinhardtii, O. tauri, O. lucimarinus, M. pusilla and M.CCMP490.*

The sequence CFATC responsible for disulfide bond formation could not be found in unicellular green algae.

Intriguingly, this is also valid for starch degrading enzymes. Glucan water dikinase (GWD), previously known as R1 protein is believed to play a key role in the process of starch mobilisation. Phosphorylating the insoluble starch granule at C-6 and C3 position, it disrupts the integrity of the insoluble granule and converts the polysaccharide to the form accessible for β-amylases and isoamylases. GWD demonstrates reversible and specific Trx regulated binding to starch granules depending on illumination state of the plant. The bound oxidized form is inactive and contains the disulfide bridge between C1004 and C1008. Chloroplast-located GWDs from potato, *Arabidopsis* and citrus share the sequence, that could not be detected in *Chlamydomonas* nor among the 5 GWD isoforms in *Ostreococcus* and *Micromonas* [92](Fig 4.2).

Another target for the redox regulation is β-amylase, the starch degrading enzyme responsible for hydrolysis of soluble starch glycans by removal of maltose. The activity of one of the plastidic isoforms of *Arabidopsis* β-amylase, TR-BAMY, has been observed to be regulated by the cell redox state. The inactive form has a folded structure with the disulfide bridge between the Cys23 and Cys 470 [93]. Again, neither *Chlamydomonas*, nor *Ostreococcus* or *Micromonas* possess the analogues of this redox-sensitive isoform (Fig 4.3).

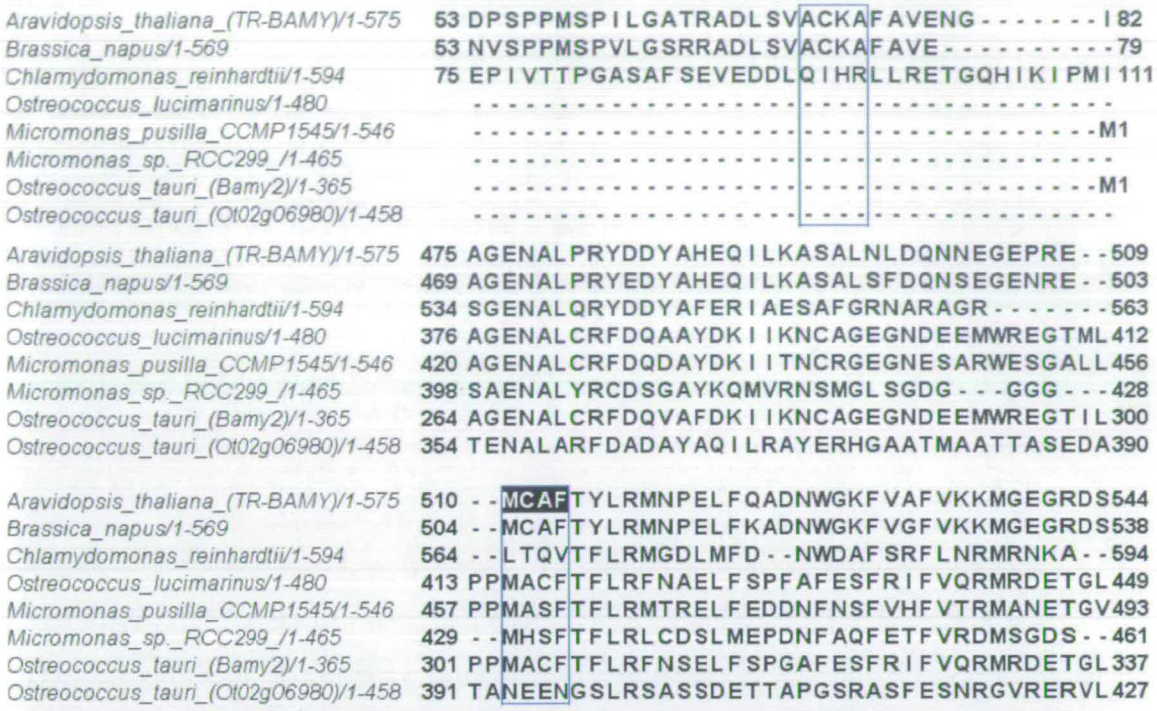


Figure 4.3 Multiple alignment for β-amylases from *A. thaliana, S. tuberosum, C. reinhardtii, O. tauri, O. lucimarinus, M. pusilla and M.CCMP490.*

The sequence MCAF responsible for disulfide bond formation could not be found in unicellular green algae.

Results for transcript and protein abundance for the *Chlamydomonas* AGPase from Ral et al [100] suggest the increased importance of regulation at the transcriptional level, namely through the circadian clock control, that contributes to physiological adjustments in unicellular algae, rather than redox regulation. It has been shown that the 24 hour period rhythm in the *C. reinhardtii* AGPase mRNA abundance is followed to a pronounced extent by enzyme activity and, consequently, by the starch content rhythm [165]. This assumption is rational for

Ostreococcus as well, as its starch enzymes, listed above, lack the redox regulatory regions. We thus propose that the metabolic fluxes in these organisms are driven to a great extent by transcriptional regulation and could be directly approximated from gene expression data sets.

4.2. Model building and structure

We started from the exploration of the Kyoto Encyclopedia of Genes and Genomes databases (KEGG) to obtain the necessary information about the proteins that comprise the starch pathway. Information in KEGG is available for *O.lucimarinus* and is partially incomplete. Nevertheless, we were able to identify orthologous proteins for all the 17 enzymes and 42 genes involved in the pathway, based on the results of the EST transcriptome analysis from [99] and BLAST sequence similarity search analysis (Table D.1).

To make the model's boundaries more specific we included the chloroplast-located gluconeogeneic subsystem, to embrace all the events for polysaccharide storage from the point of release of PGA from the Calvin cycle. The model represents the direct pathway for carbon starting from CO_2 fixed in photosynthesis, through starch granule formation, to glucose and maltose obtained from starch degradation. As we restricted ourselves to one compartment, the model includes only those reactions that are considered chloroplast specific, and transport to and from the cytosol (in the form of exchange reactions). The process of starch degradation is described up to the departure of glucose and maltose to the cytosol. (Fig D.1)

The model of the starch metabolic pathway comprises all the steps known for the higher plants and green algae [81, 82]. It starts with the conversion of glucose 1-phosphate (G1P) to ADP-glucose by means of ADP-glucose phosphorylase (2.7.7 27)[86]. ADP-glucose is used as a building block for the extension of linear glucan chains by both the soluble starch synthase (2.4.1.21) and granule-bound starch synthase (2.4.1.242)[82, 166, 167]. The starch branching enzyme (2.4.1.18) acts on the linear glucans producing branched glucans[168]. The defined ratio of long linear and highly branched glucans gives the starch granule content.

Starch degradation is initiated by starch granule phosphorylation carried out by alpha-glucan water dikinases (GWD, 2.7.9.4) [92]. Phosphorylated starch molecules can be debranched by the starch debranching enzymes (SBE, 3.2.1.68) and pullulanase (3.2.1.242), and hydrolyzed by α-amylase (3.2.1.1)

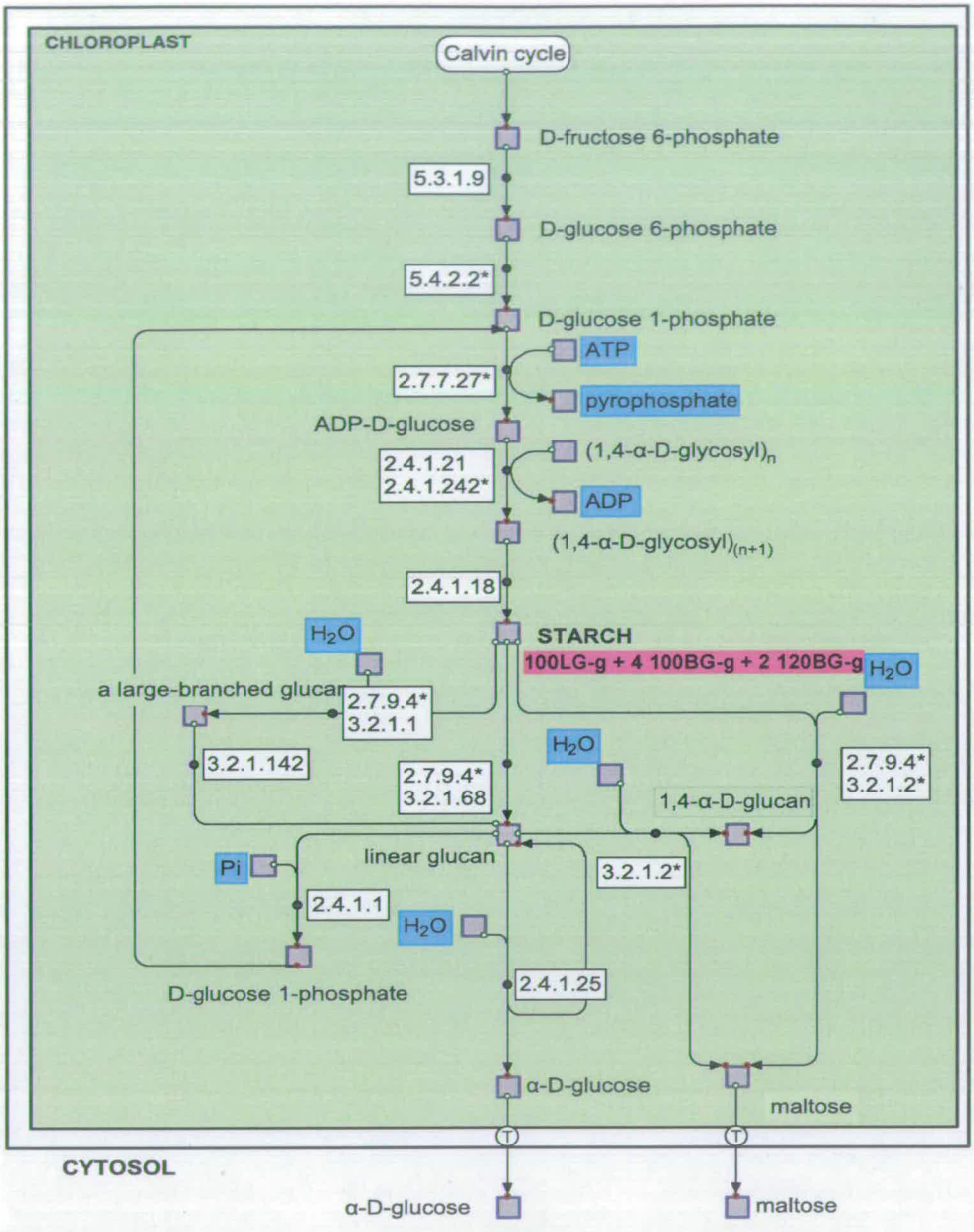


Figure 4.4 Starch metabolic pathway localized to chloroplast.

EC numbers correspond to respective enzymes: 5.3.1.9 – Glucose 6-phosphate isomerise (Ot11g02980); 5.4.2.2 - Phosphoglucomutase (Ot15g02630); 2.7.7.27 - ADP-glucose pyrophosphorylase (Ot07g03280, Ot07g02930, Ot20g00490); 2.4.1.21 – Starch synthase (Ot06g03410, Ot13g01250, Ot16g02790, Ot13g01230, Ot16g01560); 2.4.1.242- Granule bound starch synthase (GBSSI)(Ot06g03200); 2.4.1.18- Starch branching enzyme (Ot03g00840, Ot04g04110); 2.7.9.4- α-glucan, water dikinase (GWD) (Ot13g01510, Ot16g02370, Ot04g04170, Ot08g01260), 3.2.1.1 – α-amylase (Ot16g00380, Ot02g05490, Ot10g00260, Ot07g02010); 3.2.1.2 – β- amylase (Ot02g06980, Ot03g03190, Ot03g03170); 3.2.1.68- Isoamylase (Ot14g02550, Ot12g00310); 3.2.1.142- Pullulanase (Ot01g03030); MEX1- maltose transporter (Ot09g03160). Asterisks* point the potential targets for the genetic regulation revealed during the simulation.

giving short linear glucans, and by β-amylase (3.2.1.2), producing maltose and maltotrioses [166]. The latter could be disproportionated to glucose and short glucan by means of disproportionating enzyme (DPE, 2.4.1.25) [166]. Long linear glucans are also subject to the action of starch phosphorylase (2.4.1.1) giving G1P and short linear glucans. Maltose, the main product of starch degradation, and glucose are considered as the only products of starch mobilization that leave the chloroplast; their release occurs through specific transporters [169, 170].

Although the overall starch metabolism constitutes a multistep polymerisation process, it is still possible to simplify the complexity by reducing the number of metabolites to a presentable minimum, enabling one to illustrate the most essential biochemical features (Fig 4.4 and Fig D.1). To our knowledge this is the first detailed model of its kind for polysaccharide storage metabolism.

The model was developed in COPASI, converted to SBML format and analysed with the COBRA 1.3.3 Toolbox for Matlab [171]. This software allows the simulation of maximum starch accumulation using flux-balance analysis (FBA). Given the CO_2 uptake rate and the defined starch composition, the maximum possible rate of starch synthesis can be predicted *in silico*. FBA is based on linear optimization of the objective function.

The model of starch metabolism built here contains 37 metabolites and 68 reactions (Table D.3), mediated through 17 enzymes (42 genes). The key model assumptions are listed below:

- We consider all the processes described in the model to take place in one compartment (chloroplast) with a limited flux exchange with the cytosol (GAP, ATP, ADP, P, NADP, and NADPH)
- We consider starch to be produced under the light conditions from the D-glyceraldehyde 3-phosphate (GAP) obtained by CO_2 fixation during photosynthesis
- Not all the GAP produced during the photosynthesis is utilized by the starch biosynthesis, but about the half of the overall flow. The rest is

released to the cytosol, so that for every 100 molecules of CO_2, giving about 33 molecules of GAP, 16 of them go to the cytosol and 17 to starch

- Starch consists of amylose and amylopectin in the approximate ratio 1:9 [81]. Our model starch consists of long linear glucans (100 glucose residues) and large branched glucans (100 and 120 glucose residues), thus we consider starch as a composite of three main components: 100LG, 100BG and 120 BG in the ratio 1:2:4

  The exchange reaction for starch formation in our model, thus, has the form of: **100LG-g + 4 100BG-g + 2 120BG-g** <=>, that fixes the ratio ('starch_ex')

- We also take into consideration short linear glucans (20 glucoses), middle linear (40, 60 and 80 glucoses) and middle branched (60 and 80 glucoses). All these species work as intermediate steps during the branching and debranching processes
- The branching happens without elongation of the chain, just by the rearrangement of the existing linear chain by SBE (2.4.1.18)[168]. Elongation happens by means of the SS (2.4.1.21) and GBSSI (2.4.1.242) enzymes with ADP-glucose (ADPG) as a building block [100]
- Both small glucans (20LG) and glucose-1-phosphate (G1P) are allowed reutilization during the starch synthesis as building components; thus we get the number of internal cycles within the whole pathway
- We assume that the components of the starch granule can be accessed by degradation enzymes only after phosphorylation by alpha-glucan, water dikinase enzyme (GWD) [92]
- Only the components of the granule surface are accessible for phosphorylation followed by subsequent degradation
- We consider the starch granule as a separate compartment. We design an exchange of surface granule components with the granule. Once starch components come to the granule and are covered by a new layer of surface components they are no longer available to degradation enzymes (Table D.3).

4.3. Extreme pathway analysis

To study the basic network properties, we started with the exploration of all possible biochemically meaningful flux solutions known as extreme pathways, since they correspond to the null space boundaries [4]. As the number of extreme pathways tends to grow faster than the number of components of the network, our system predictably gives a large number of extreme pathways, namely 24000.

The extreme pathways are classified based on the exchange reactions that they include:

- Type I pathways contain the external exchanges and correspond to conversion of primary inputs to primary outputs. Our system has 167 pathways of that type
- Type II pathways contain irreversible internal exchanges of currency metabolites, so-called futile cycles. Our system involves 275 pathway of this type (Fig 4.5, A). We also calculated the distribution of the pathway lengths for the pathways of this type (Fig 4.6). It is clear from the picture that the majority of pathways have a length of 28-32 reactions
- Type III pathways contain internal cycles with zero net flux. Our system has five elementary pathways of this type: 100LG_g ⇔, 120BG_g ⇔ , GAPiso ⇔, starch_ex ⇔, 100BG_g ⇔

We calculated the reaction participation values to identify the reactions with high and low participation (Fig 4.5, B).

Participation above 90% is characteristic of the following reactions (suffixes 'b' and 'f' correspond to back and forward reactions, respectively, that were considered separately for the reversible reactions) : 'PPitoP', 'ADPG', '100LG g_b ',' 100BG g_b, 120BG g_b', 'starch ex_b', '100LG s P', '100BG s P', '120BG s P', 'ATPex_f', 'ADPout', 'Pex_f', and 'maltose out' (Fig D.1).

The participation value of ATP and ADP exchange is reasonably large since the system is driven by the external supply of ATP. On the other hand, the reactions

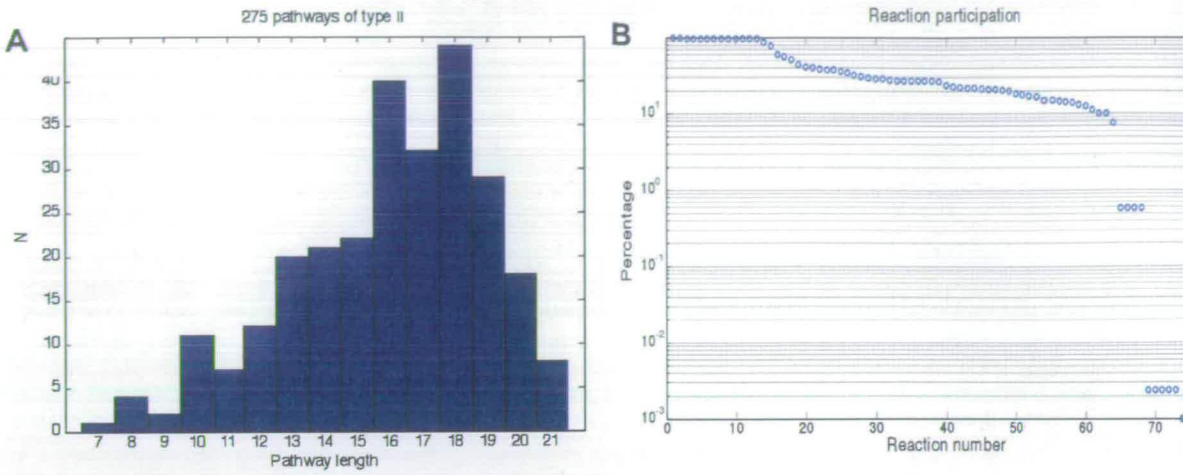


Figure 4.5 Results of extreme pathway analysis
A. The distribution of pathway lengths between the 275 pathways of type II
B. Reaction participation presented on log-log scale

corresponding to initiation of starch degradation also have a comparably high participation

There are ten reactions ('GAPphoto', 'GAPiso_b', '100LG g_f', '100BG g_f', '120BG g_f', 'starch ex_f', '20LGfromADPG', 'CO_2_r', 'NADPHex_f', 'NADP_f') with participation below 0.01%. A low participation value indicates those reactions, which have highly specific network functions.

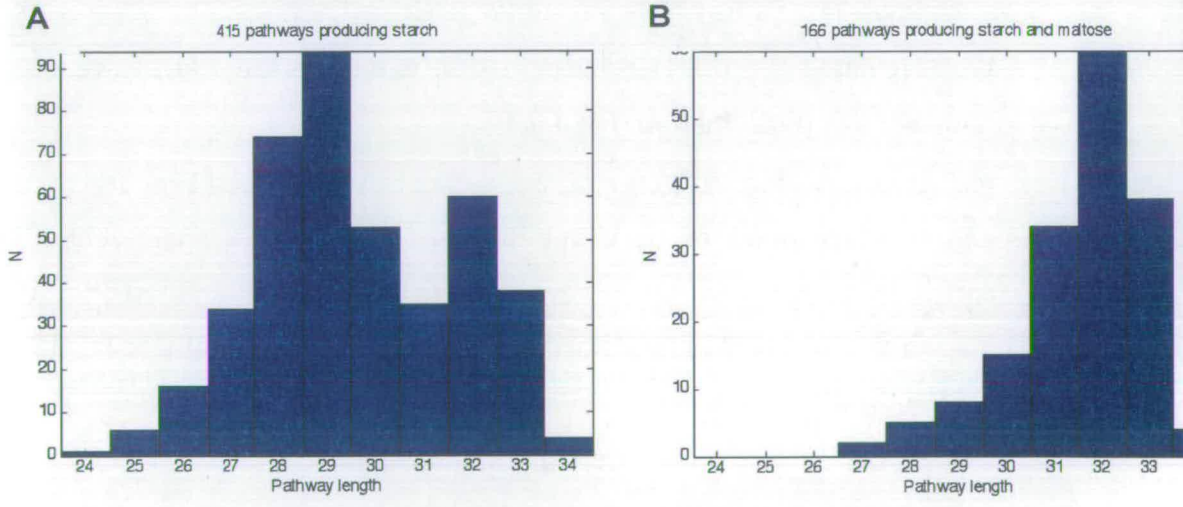


Figure 4.6 The distribution of pathway length for starch and maltose production.
A. Pathways producing starch B. Pathways producing maltose

Estimated CoSets (the set of reactions that usually appear correlated in analysis) give the information about the hierarchical structure of the network and have to be borne in mind during the experimental design. We identified the 21 reactions CoSet related to starch biosynthesis that appears in 415 pathways: 'starch ex_f', 'GAPiso_f', 'F16BP', 'F6P', 'G6P', 'G1P', 'PPitoP', 'ADPG', '60LGfrom40LG', '80LGfrom60LG', '100LGfrom80LG', '120BGfrom100BG', '100LG s', '100LG g_f', '100BG s', '100BG g_f', 120BG s', '120BG g_f', 'GAPex_f', 'ATPex_f', 'ADPout', 'Pex_f (Fig D.1). The CoSet comprises the linear part of the pathway that mainly consists of the gluconeogenic reactions. Since this linear path is an actual bottleneck for both the model and real systems, it works as a rate-limiting section providing the only input of glucose (Fig 4.4 and Fig D.1).

We also identified a unique CO_2 consumption extreme pathway which contains 8 reactions: 'GAPphoto', 'CO_2_r', 'GAPex_f', 'NADPHex_f', 'NADP_f', 'ATPex_f', 'ADPout', 'Pex_f'.

4.4. FBA

We analyse the dynamics of the starch metabolite during the light/dark cycle as the sequence of metabolic quasi-steady states, each of them corresponding to particular genetic state within the 3 hours interval. We developed the detailed reaction model of starch metabolism which, to our knowledge, is the first attempt to describe the polysaccharide polymerization preserving the mass balance relationships (generics like $(1, 4\text{-glucan})_{n+1}$ common for this pathway representation do not allow mass –balance modelling). Previous reaction models for carbohydrate metabolism usually present starch synthesis as ADP-glucose turns to starch without intermediates. Although being useful simplification when the whole organism metabolism is modelled, it doesn't allow the incorporation of full gene essentiality. That is why we were trying to keep the degree of graining as large as possible for the beginning, though it might be reduced for the future work.

Since a starch polysaccharide has an enormous molecular mass, the normal flux value for starch production will be in the range of 10^{-5}—10^{-6} mM/h, which is

difficult to analyse, as it is prone to rounding errors. To avoid this, we set all the boundaries for our model reactions as a proportion of maximum starch production rate. This enables us to operate with solutions in which most of the flux values are greater than one.

The stoichiometric matrix for the system is of size 38 x 68. The vector $\boldsymbol{v}$ has 68 fluxes, including 6 exchange fluxes (b1-b6):

$\boldsymbol{v}$ = [v1 v2 ... v62 b1 b2... b6]^T

Number of approaches for utilisation of expression activity information in constraint-based modelling was published recently [163, 164]. Shlomi et al. proposed method to take into account tissue specificity in metabolic models by incorporating expression states into flux optimisation scheme [163]. Their method allows combining determination of FBA solution together with maximisation of correspondence between expression status and metabolic fluxes. At the same time representation of expression level with tri-valued method is significant simplification and leads to destruction of dynamical pattern especially when more than one isozyme is responsible for reaction regulation (data not shown). As we are trying to apply the expression time series to analysis of dynamic of metabolic system, we apply the variation of each gene expression in diurnal cycle as a determinant for maximum flux constraint, modifying the corresponding reaction upper bound. The way we are doing this resembles the E-flux method where the flux upper constraints are represented as a function of measured gene expression [164], but in our case modifications of the flux upper bounds are performed in a slightly different manner.

Objective function

We used linear optimization [4, 171] for finding solutions of interest within bounded null space, defined by:

$$S_{exch}\begin{pmatrix} v \\ b \end{pmatrix} = 0, \text{ where } 0 \le v_i \le v_{i,\max}, \text{ and } b_{i,\min} \le b_i \le b_{i,\max}$$

A linear objective function to define has a form:

$$Z = w * \begin{pmatrix} v \\ b \end{pmatrix} = \sum_i w_i v_i + \sum_j w_j b_j,$$

where **w** is a vector of weights (w_i) on the internal flux (v_i) and the external flux (b_j). **Z** then has to be optimized to minimum or maximum, depending on the problem posed. The constrained optimization procedure is known as *linear programming* (LP).

First, we introduced light into the system by equating the weight coefficient for the CO_2 uptake flux in the objective function to one. As a result, nightfall could be easily simulated by switching it to zero, assuming saturating light.

There is no data about dependence of starch production and sugars exchage in autotrophic growth of *O. tauri* available to the best of my knowledge, so for applying FBA to the system I have developed cost function from first principles and simple biological intuition.

In the initial setup when the cost function includes only the starch production, the model could either produce starch (if optimized to maximum), or maltose (optimized to minimum). A competing relationship between starch production and maltose export for carbon fixed by the photosynthetic machinery makes net day production rate negative, which contradict experimental observation as complete lost of starch was never observed. It is easy to understand such behaviour as the path from ADP-glucose to maltose is much shorter and less regulated compare to starch production, so we have to introduce the specific component to cost function for balansing maltose and glucose transport by starch production.

Among the expression data we have information about gene activity of maltose transporter. However, the direct application of gene regulatory constraints to infinite flux through maltose transporter is unreasonable. Since the experimental data about its exact transport capacity is not available, we need to estimate maltose and glucose efflux upper bounds.

To estimate five missing values (three coefficients in cost function and two flux upper bounds) we decides to use two majour biological intuitions: (i) in

autotrophic conditions the only source of carbon is photosynthesis and it existst during the light period only and (ii) net starch production is positive. Since we aim to model the diurnal cycle in starch production and degradation, and having in mind that starch is generally produced during the day when the carbon source is available, we asumed that the starch production rate is somehow similar to the step function (+1 during the day and -1 at night). The step function starch production was fitted by simultaneous modification of the maltose transport flux upper bound and the weight coefficient for the corresponding components in the cost function. The optimum has been achieved with the values for the export reactions upper bounds equal to 166.5 and 122.45 for maltose and glucose respectively, which is slightly less than the stoichiometric capacity of the system and corresponds to positive starch production rate.

The resulting general linear objective function includes the CO_2 fixation, starch production, and glucose and maltose export from the chloroplast:

$$Z = 1 * CO2 + 10468.135 * starch_{ex} + 963.325 * maltose_{out} + 650 * glu_{out}$$

The objective function presented above is in fact a heuristic function selected on the basis of its ability to produce biologically meaningful form of output for the stach production. It will require validation by experimental measurements of specific fluxes although, in part, it was supported by a good similarity between predicted and experimentally measured dynamic profiles of starch content.

The optimal flux distribution for the objective is presented below. Only reactions with flux values different from zero are depicted at Figure 4.7.

As the quantitative metabolic data for biochemical status for O.tauri is sparse we have to develop our own scheme to represent the results of analysis. We normalize all flux values acoording to fluxes in ideal system where network utilize all its capacity and rate of starch production (flux through 'starch_ex' reaction referred thereafter as Starch flux) has a unit value. Relative flux value will be referred elsevhere as a relative flux unit (RFU).

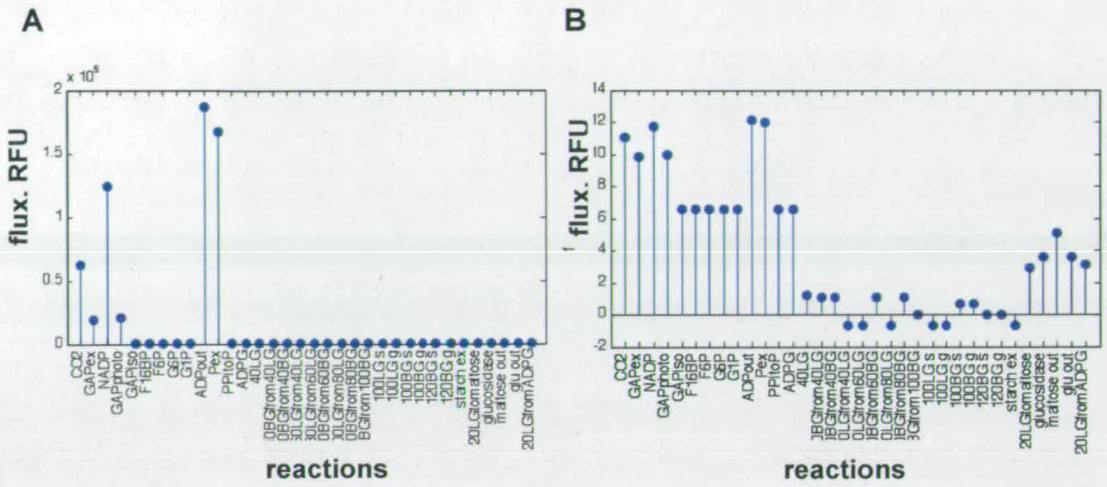


Figure 4.7 The optimal flux distribution for the objective function
(A) The $\log_{10}$ representation is more informative as the distribution of flux values covers several orders of magnitude (B)

Flux variability analysis

We determined the range of fluxes that correspond to the optimum solution by means of Flux Variability Analysis (FVA), provided by the COBRA Toolbox [170]. The range has been computed for each flux by solving the following LP problems [4, 171]:

maximize or minimize v_i

subject to: $Z = Z_{opt}$

$\mathbf{S^*v = 0}$

$v_{i,\min} \le v_i \le v_{i,\max}$ for i=1,..., n

The results of FVA are presented on Figure E.1. It is clear from the figure that due to the given objective function those fluxes contributing to the objective ('CO_2', 'starch_ex', 'maltose_out', 'glucose out') have a fixed value and, accordingly, zero range. The same is true for the linear gluconeogenic and ADP-glucose producing branch of the pathway ('F16BP', 'F6P', 'G6P', 'G1P', 'and ADPG'), as it is strictly constrained by the objective function for a starch synthesis rate set to maximum. Flux values for other components vary

depending on how compensated this particular reaction is relative to the whole pathway. For instance, for reactions from the debranching part of the model the range could be noticeably larger than for reactions from the branching part because the combinatorics of debranching options allows substitution of one reaction by another without a loss of the overall pathway productivity.

Robustness analysis:

COBRA Toolbox allows us to compute the effect of varying flux through a single reaction. The objective function changes as a flux through the particular reaction varies in magnitude, giving an idea of sensitivity and the overall importance of each reaction relative to the starch production[4, 171](Fig E.2).

We can group all the reactions into five classes according to their influence on the starch production rate: linear positive, linear negative, non-sensitive, one-side low rate plateau and one-side high rate plateau.

For each reaction of the linear starch biosynthetic pathway we can observe the same linear positive relationship with the starch production rate (Fig E.2: reactions 6, 7, 8, 9, 10). This property emerges from the model and pathway structure (these reactions comprise the linear part of starch the CoSet mentioned above (Fig D.1)) as these reactions directly define the rate of starch biosynthesis. We determine these reactions as extensively sensitive to small perturbations, and thus as attractive potential regulatory targets.

Reactions involved in branching demonstrate either positive (Fig. E.2: reactions 21, 25, 28, 29, 31, 33) or negative (Fig E.2 reactions 17, 18, 23, 24, 27) flattening behaviour depending on whether they increase or decrease the starch production rate. Although both contribute to starch granule formation, the latter divert the precursors of the main starch components from the direct course. The starch synthesis rate is sustained near the optimum value for these reactions, indicating that the network is robust with respect to their flux changes. It is not surprising, for they comprise the networked structure of combinatorial events giving the same result: starch granule formation.

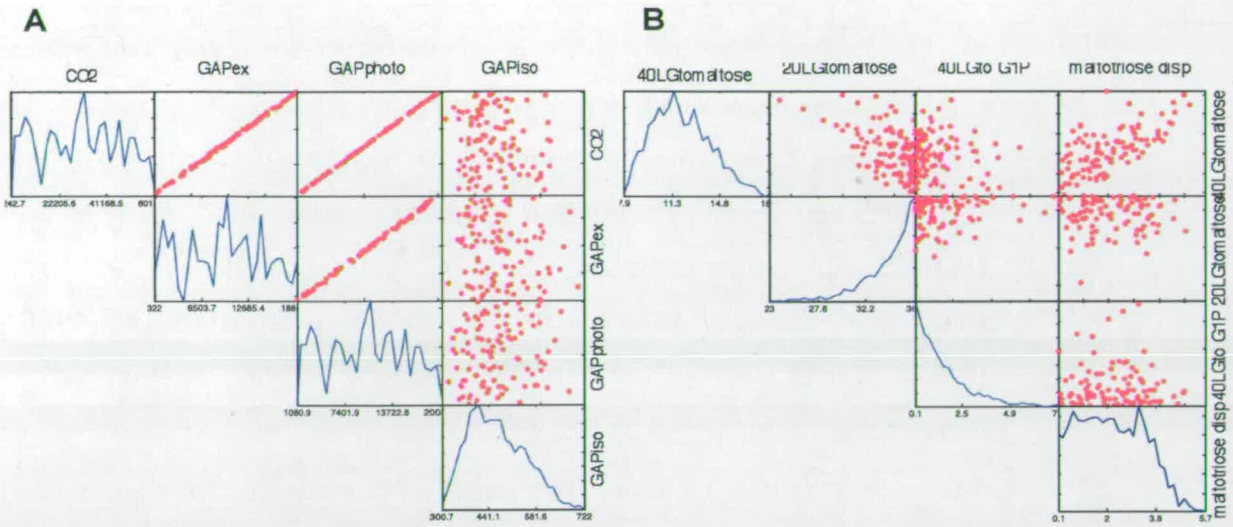


Figure 4.8 Some results of sampling flux distribution.

Flux distribution histograms (diagonal) and pairwise scatterplots (off diagonal) for 4 initial and 4 final reactins of the model pathway (reaction names are the same as in Fig G.1) X axis corresponds to the magnitude of the flux through the particular reaction. The off- diagonal scatterplots show the relationships between fluxes through two reactions.

All the reactions participating in starch degradation give low-level plateau pattern as at a certain flux value they are responsible for diminution of the starch pool. As in the previous case, they describe the multichoice mechanism of the complex glucan decomposition. As a result, the system as a whole is less sensitive to each of them. However, their importance for the starch degradation correlates with the rate of the plateau. Some of them, such as reactions 41, 42, 47, 48, 50 (α-amylase and isoamylase) look to some degree accessible for regulation.

Finally, the reactions for β-amylase and starch phosphorylase provide the sink of the system: the conversion of middle and short glucans to maltose and glucose. Therefore they demonstrate the highly sensitive, linear negative pattern and, similar to the synthetic reactions, appear to be a good target for regulation.

Sampling of solution space

We used the COBRA Toolbox sampling of flux distributions that are allowed by stoichiometric properties of the network and given flux capacity constraints. The information obtained from the analysis is useful in respect of predicting the

most likely flux value through the particular reaction, as well as the dependence relationships between reaction pairs. For instance, CO_2 uptake is highly correlated with reactions GAPphoto, which accounts for the photosynthesis, and GAP exchange, corresponding to the release of the half of the GAP to the cytosol (Fig 4.8, A). At the same time it is independent from the reaction of GAP isomerisation, and also the reaction for maltose production from 40LG glucans (40LGto maltose) is rather independent from reaction of maltose disproportionating (Fig 4.8., B).

To summarize, the stoichiometric analysis presented above gives the general representation of the capacity of the system and potential variability of its behaviour, depending on applied constraints. Particularly, the extreme pathway analysis sheds light on how the system may behave if it not constrained by the objective function. The sampling of the solution space and, especially, the robustness analysis are extremely useful as they both provide the idea of each reaction's "importance" in terms of the overall system performance and uncover the existing relationships between the reactions. The latter should be taken into account in experimental design.

4.5. Investigation of timing

All reactions in our model are gene associated. This means that each enzyme responsible for a certain reaction has one or several genes of which the function has been identified correspondingly. Each gene has a characteristic pattern of diurnal expression that could be correlated to its specific physiological function. On average, as one would expect, the peak expression of the majority of starch biosynthetic genes occurs during the day. On the contrary, genes corresponding to starch degrading enzymes have peak expression in the night. Nevertheless, there is no strict relationship between the peak of expression and the proposed function. For instance, some isoamylases and α-amylases are expressed during the day while some of the starch synthases are preferentially expressed during the night (Fig 4.9). In the light of this variability it was interesting to look at the combined effect of all genes on the diurnal starch pattern.

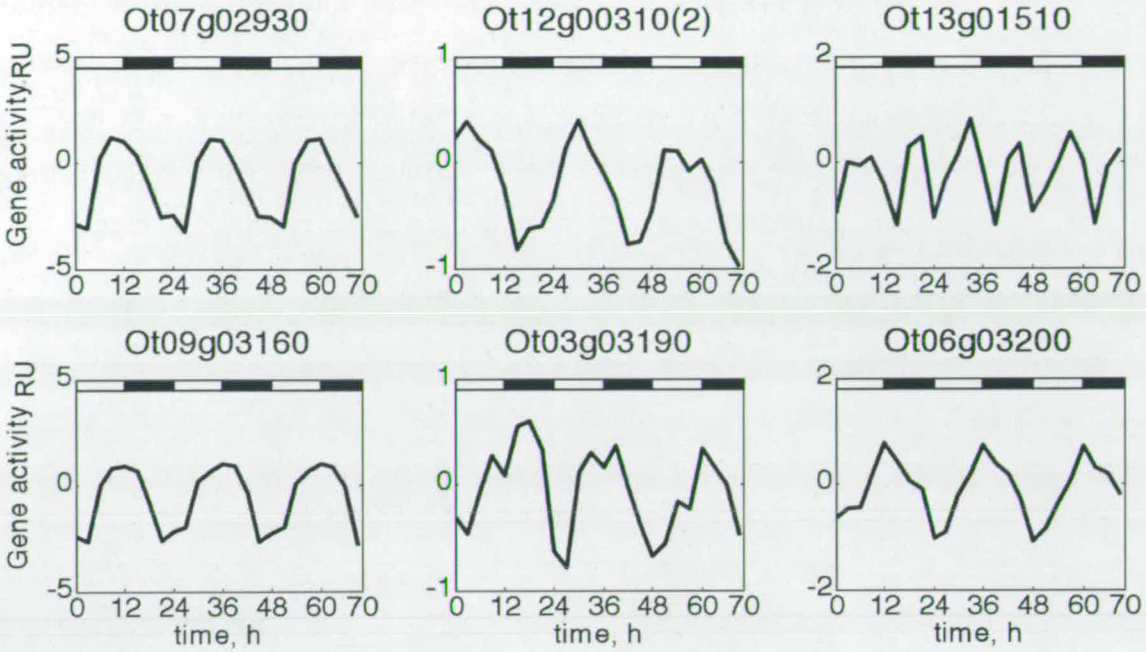


Figure 4.9 Diurnal expression profiles for selected genes

From left to right upper row: ADP-glucose pyrophosphorylase, (Ot07g02930), isoamylase ISA3 (Ot12g0310), GWD(SEX1) (Ot13g01510); Lower row: maltose transporter MEX1 (Ot03g03160), β-amylase (Ot03g03190), GBSSI (Ot06g03200).

Application of microarray data to flux analysis

Variation of gene expression could lead to deviation of reaction flux upper bounds from their value defined by the stoichiometry of the system. To represent such a deviation we scaled RNA abundance from the microarray data to range from 0 to 1 and multiplied the flux upper bound by the value obtained. Thus, we represent dynamics of starch content as a set of steady states corresponding to genetic states available through microarray data. Each steady state corresponds to a certain time point of the 12/12-light/dark cycle with an interval of 3h between the neighbouring points. Data are provided by François-Yves Bouget (personal correspondence).

4.5.1. Investigation of each gene's relative impact on the overall starch production

We started from the investigation of each gene's relative impact on the overall production of starch. In other words we are looking at how the starch flux rhythm responds to the rhythmical expression of each gene. For this we apply

each gene's microarray data, one by one, multiplying the corresponding model reaction upper bound by the microarray value. Introducing a certain gene expression profile we simulate the situation when all the genes, except the chosen one, are expressed at their maximum level. Thus, we interrogate how the flux limitation obtained from single gene regulation affects the overall starch production flux (Fig 4.10).

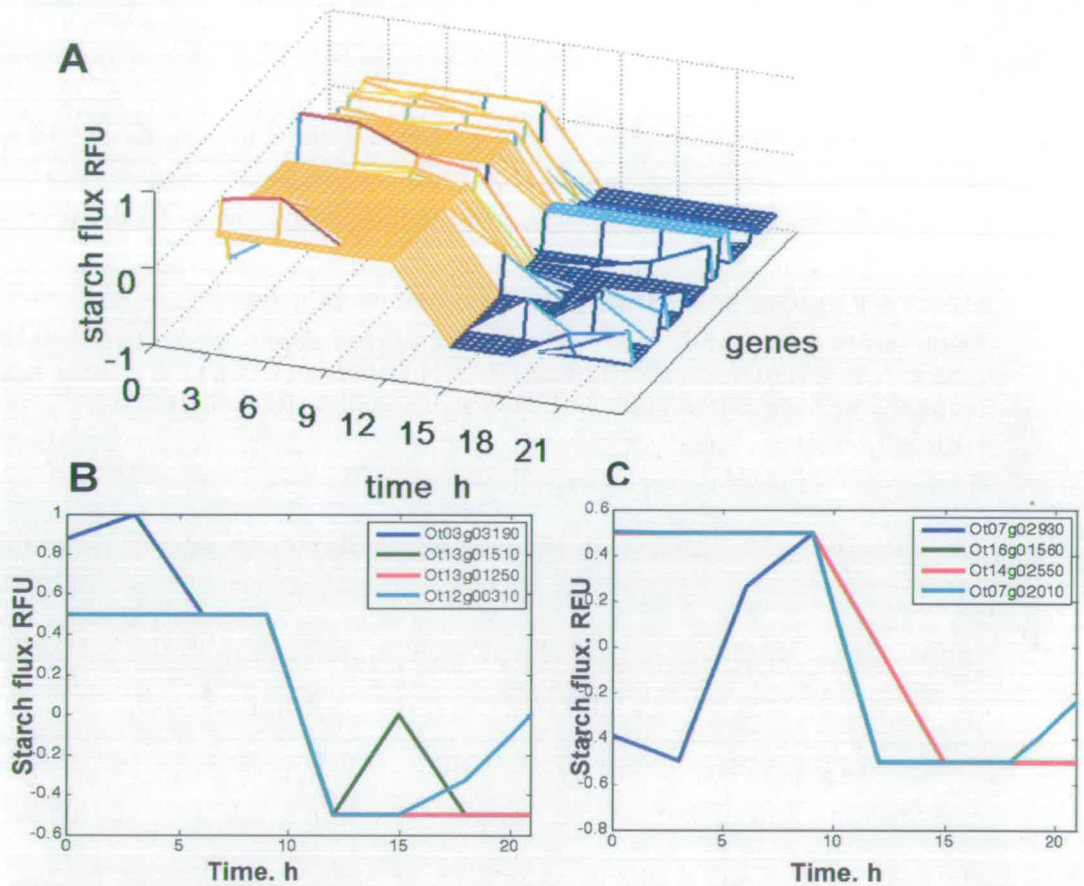


Figure 4.10 The investigation of single gene rhythmic expression in the overall diurnal starch pattern.

Time 0 corresponds to the light onset, time 12 - to the dark onset. Flux through 'starch_ex' exchange reaction shoawn as Starch flux. A. A 3D representation of each gene's impact. Shown is the hypothesis that all genes except one are expressed at their maximum level. Note time pointes 1 to 8 shown from right to left for the purposes of visualization. B. The effect of expression pattern of Ot03g03190 (β-amylase), Ot03g01510 (glucan water dikinase) and Ot12g00310 (isoamylase) on the overall starch shape. Ot13g01250 profile is shown for comparison and corresponds to the unperturbed shape of the step-function. C. The effect of Ot07g02930 (AGPase), Ot14g02550 (isoamylase). Ot7g02010 and Ot16g01560 correspond to the unperturbed step-function.

We revealed that Ot03g03190 (β-amylase) and Ot07g02930 (AGPase), strongly affect the 'light' part of starch flux pattern, while Ot03g01510 (glucan, water dikinase), Ot12g00310 and Ot14g02550 (isoamylases), Ot07g02010 (α-amylase) have an influence on the 'dark' region.

4.5.2 Investigation of combined influence of all the genes on the starch diurnal pattern

Next we interrogated the combined impact of all genes involved in the pathway. We substitute the entire given microarray data *simultaneously* to track the fluxes for starch and maltose production during the day. For this, at each time point all the reactions' upper bounds are modified in the way described above. As we have more than one gene associated with some reactions we sum up the gene data for each reaction and multiply the current upper flux bound by the obtained sum:

$$v^t_{i,max} = v^{\infty}_{i,max} \sum_j c_{i,j} \frac{\alpha^t_j - \min_t(\alpha^t_j)}{\max_t(\alpha^t_j) - \min_t(\alpha^t_j)},$$

where $v^t_{i,max}$ is reaction i upper bound at time t; $v^{\infty}_{i,max}$ stoichiometric capasity of reaction i in RFU; $c_{i,j}$ reaction-gene binding coefficient, $c_{i,j} = 1$ if gene j regulates reaction i and $c_{i,j} = 0$ otherwise; α^t_j - gene activity at time t. Proposed procedure based upon assumption that each gene in its full activity is able to carry all regulated reaction at the rate of stoichiometric capacity.

As before, we introduce the LD conditions allowing the CO_2 uptake from ZT0 until ZT12 (Fig 4.11).

We obtained the maximum starch synthesis flux at ZT6 and the peak degradation flux at ZT12. Interestingly, we observe the coinciding peak of maltose export flux at ZT12, which is biologically meaningful, as the maximum rate of maltose production must be related to the maximum rate of starch degradation. Figure 4.11 with the two fluxes plotted together illustrates the

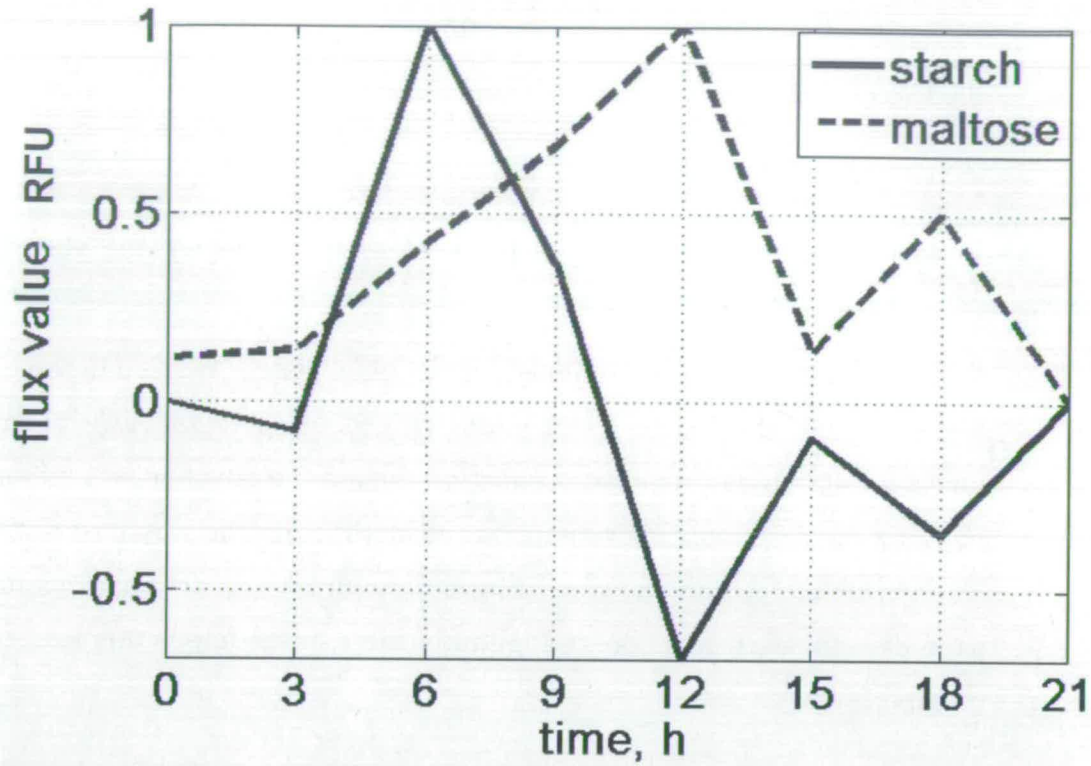


Figure 4.11 Diurnal rhythm of starch and maltose fluxes presented on the same scale.
Peak in maltose flux at ZT12 corresponds to the trough of the starch flux

negative correlation between starch and maltose fluxes. It is known for plants that the peak maltose release occurs during the middle of the night [169].

Nevertheless, the starch granule division in *Ostreococcus* coincides with cell division and happens at the end of the day [99]. It is not unlikely that these events affect the timing for the maximal starch elongation and degradation.

Figure 4.12 shows the flux distribution within each time point.

At time zero we have no starch production, and CO_2 is already available. The system converts the remaining short glucans to maltose and glucose (Fig 4.12)

At time point ZT3 we observe quite an unexpected situation. Regardless of the CO_2 uptake and ADP-glucose synthesis the system operates towards starch degradation. This reflects the combined action of gene expression data and

should be clarified with experiments, but to speculate, it may account for increased need for rapid sugar supply in the early day.

At ZT6 we see the normal starch assimilation pattern with all steps of glucan synthesis participating. The same is true for ZT9. However, note that the absolute flux value is smaller for the reactions involved in starch granule formation (100LG_s, 100LG_g, etc.).

At ZT12 the starch flux leaves the positive region, and the process of degradation is dominant.

The fluxes of the reactions for starch granule phosphorylation are brought to the forefront; meanwhile, the flux for maltose export attains its maximum value (Fig 4.13).

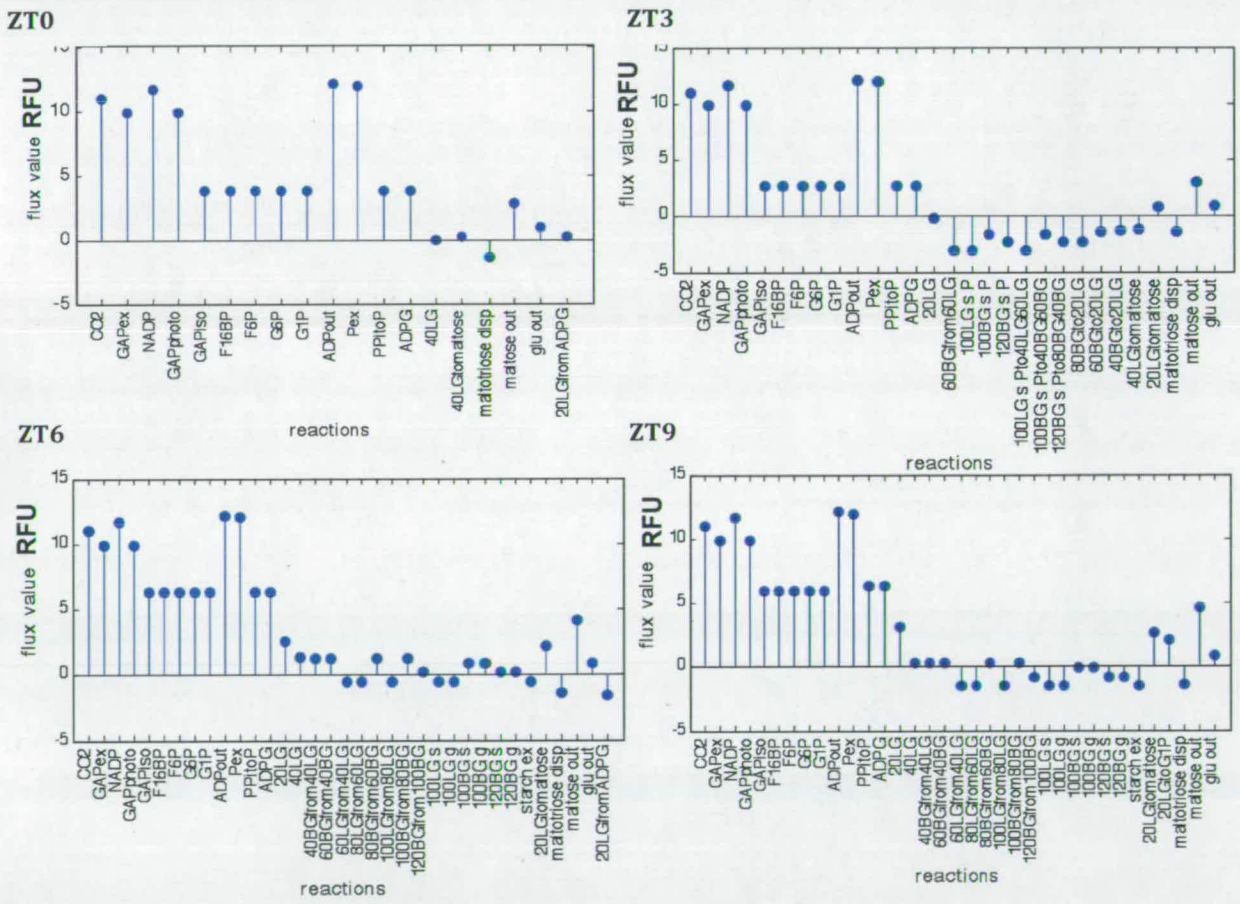


Figure 4.12 The log flux distribution at ZT0, ZT3, ZT6 and ZT9 (only non-zero fluxes are represented)

At ZT15 and ZT18 starch degradation is going on, though the fluxes values have relatively smaller values. At ZT21 all the fluxes are equal to zero (not shown). This is caused by a zero value for maltose efflux, which, in turn, results from a zero value for the expression level of MEX1. In my procedure, this is directly applied as the upper flux bound for the reaction of maltose translocation from the chloroplast, but merits further experimental investigation.

4.5.3 Investigations of the dynamics of starch content

If we plot the cumulative value of the starch flux, we obtain the diurnal dynamics of the starch content. Figure 4.14 shows the comparison of predicted pattern for the starch content and those obtained from experiment. The predicted curve matches the real starch content behaviour with a very high

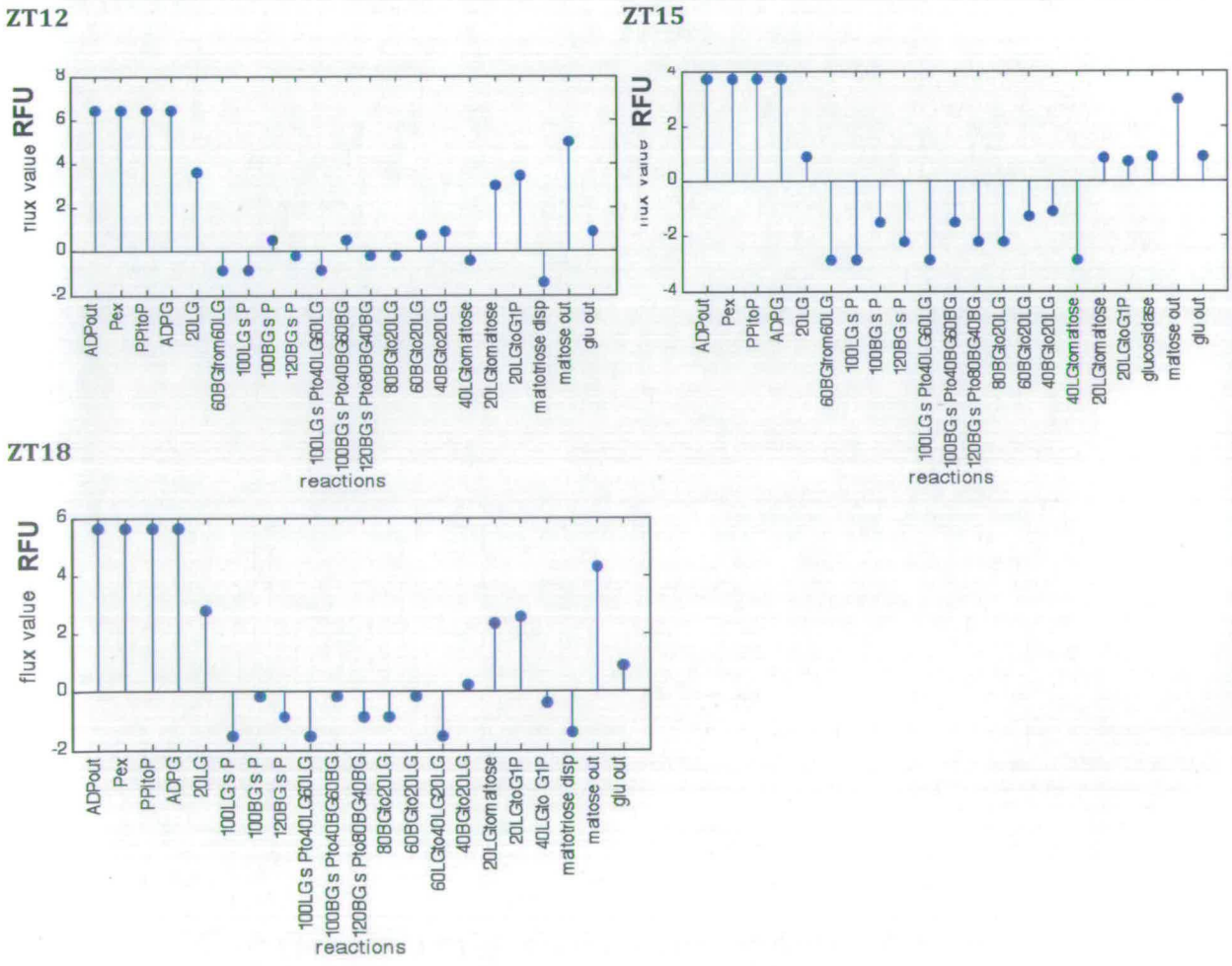


Figure 4.13 The log fluxes distribution at ZT12, ZT15 and ZT18 (only non-zero fluxes are presented)

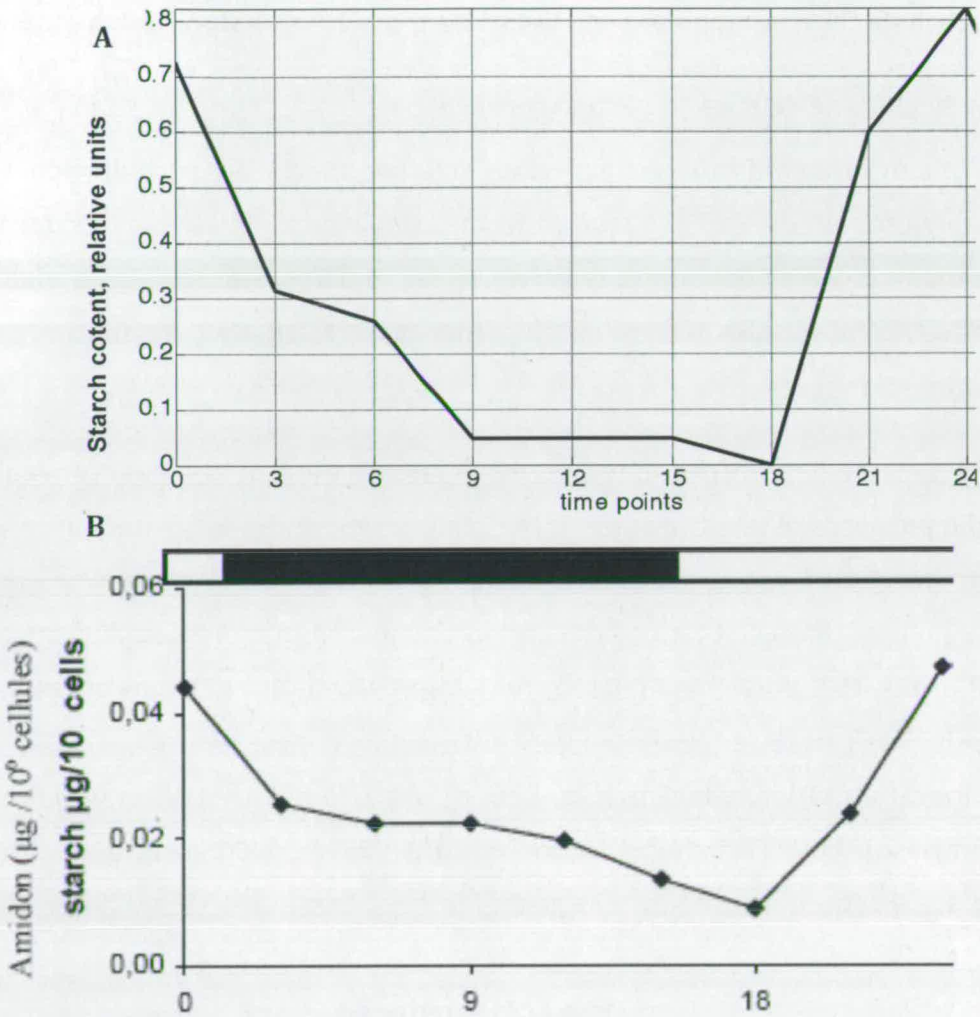


Figure 4.14 The diurnal patterns of the starch content obtained by the flux summation and in experiment

(A). The diurnal pattern of the starch content obtained from simulation
(B). The experimental data for starch content dynamics (provided by J-P Ral, Steven Ball and colleagues (unpublished), by the kind offices of François-Yves Bouget (personal communication)).

level of accuracy. Maximum content falls after about 10 hours of light. This again seems physiologically reasonable, as the cell should accumulate sufficiently high level of polysaccharide storage before the moment of its distribution between the two daughter cells (Fig 4.14).

If we simulate the weekly behaviour of starch content, we will see the tendency to increase (Fig 4.15). The overall starch production/degradation ratio is clearly displaced towards net production, as was enforced by choise of coststt function.

Given the necessity for cell volume increase and starch granule division, we can conclude that the trend toward the net starch accumulation is realistic enough.

4.5.4 Single gene deletion in silico experiment

We aim to identify the genes whose deletion affects the periodic behaviour of starch accumulation. We started from a single gene deletion experiment. Each gene microarray data point in turn was set to 0.01. After this, the simultaneous effect of the substitution of all genes was calculated as in the previous simulation (Fig 4.16).

3D representations presented here and later in the text are most convenient for the purposes of identification of the regions where the estimated diurnal shape of starch flux is perturbed. For instance, Figure 4.16, A gives three potential regions of interest, that could be further analyzed in more detail. From the first 8 genes, the pattern is affected most impressively by deletion of Ot06g03200 (GBSSI). This is a predictable result as the enzyme is responsible for the elongation of long glucans (Fig 4.16, B). Also, the 'dark' region of pattern (β–amylase); both straighten the pattern after ZT12, decreasing the degradation rate. From the middle 7 genes we obtain the Ot7g02930 (AGPase)

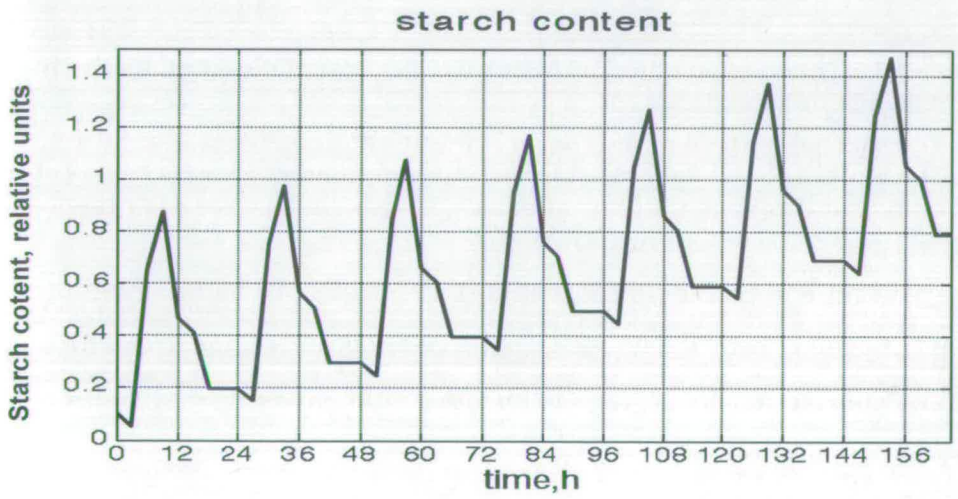


Figure 4.15 The weekly pattern of the starch content dynamics.

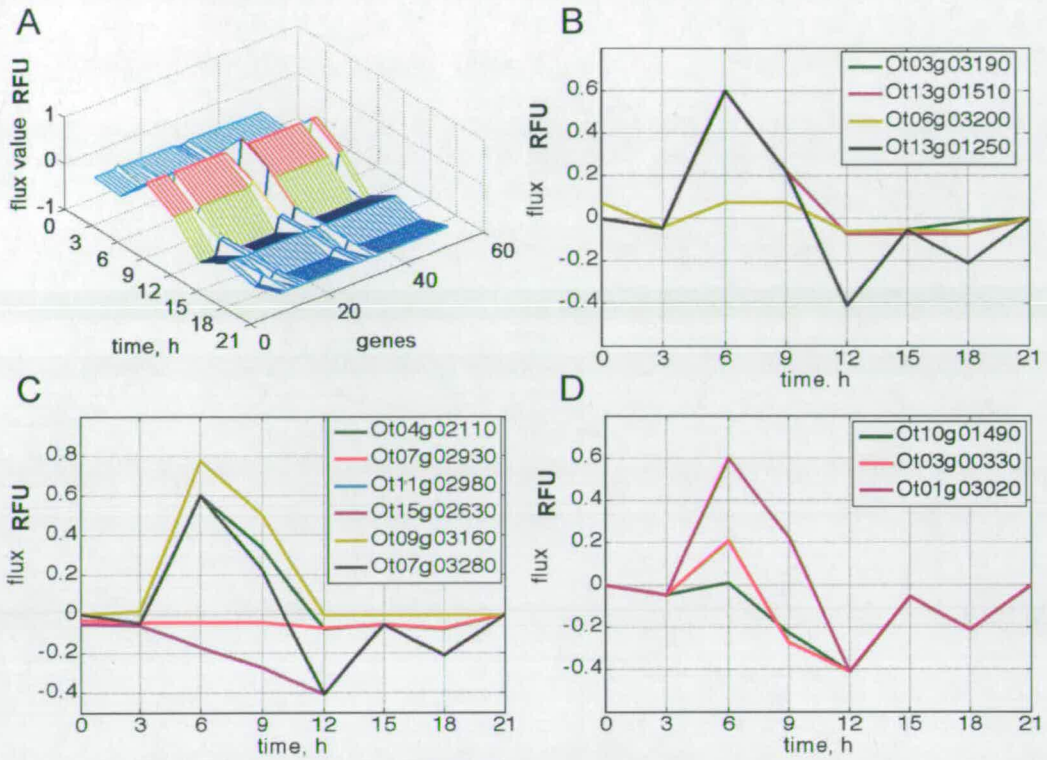


Figure 4.16 In silico single gene deletion experiment.

A. A 3D representation of the results of a single gene deletion experiment.
B. Effects of deletion of GBSSI (Ot06g03200) (GBSSI), GWD (Ot13g01510) and β-amylase (Ot03g03190) on the overall starch flux diurnal pattern. SSIII-B (Ot13g01250) corresponds to the unperturbed starch pattern.
C. Effects of deletion of AGPase (small subunit, Ot7g02930), phosphoglucomutase and glucose 6-phosphate isomerise (Ot15g02630 and Ot11g02980, overlay each ather), MEX1 (Ot09g03160), starch phosphorylase (Ot04g02110) on the overall starch flux diurnal pattern. AGPase, large subunit (Ot07g03280) corresponds to the unperturbed starch pattern.
D. Effects of the deletion of the fructose-1,6-bisphosphatase, fbaII (Ot03g00330) and fructose-bisphosphate aldolase (Ot10g01490) on the overall starch flux diurnal pattern. Fructose-1,6-bisphosphatase, fbaI (Ot01g03020) corresponds to the unperturbed starch pattern.

that removes the oscillatory pattern completely when deleted (Fig 4.16, C). This is again the main rate-limiting step of the biosynthetic pathway and is a potentially very sensitive regulatory target. The phosphoglucomutase, Ot15g02630, and G6P isomerase deletion moves the starch flux to the negative region, towards degradation. The results for AGPase, phosphoglucomutase and G6P-isomerase

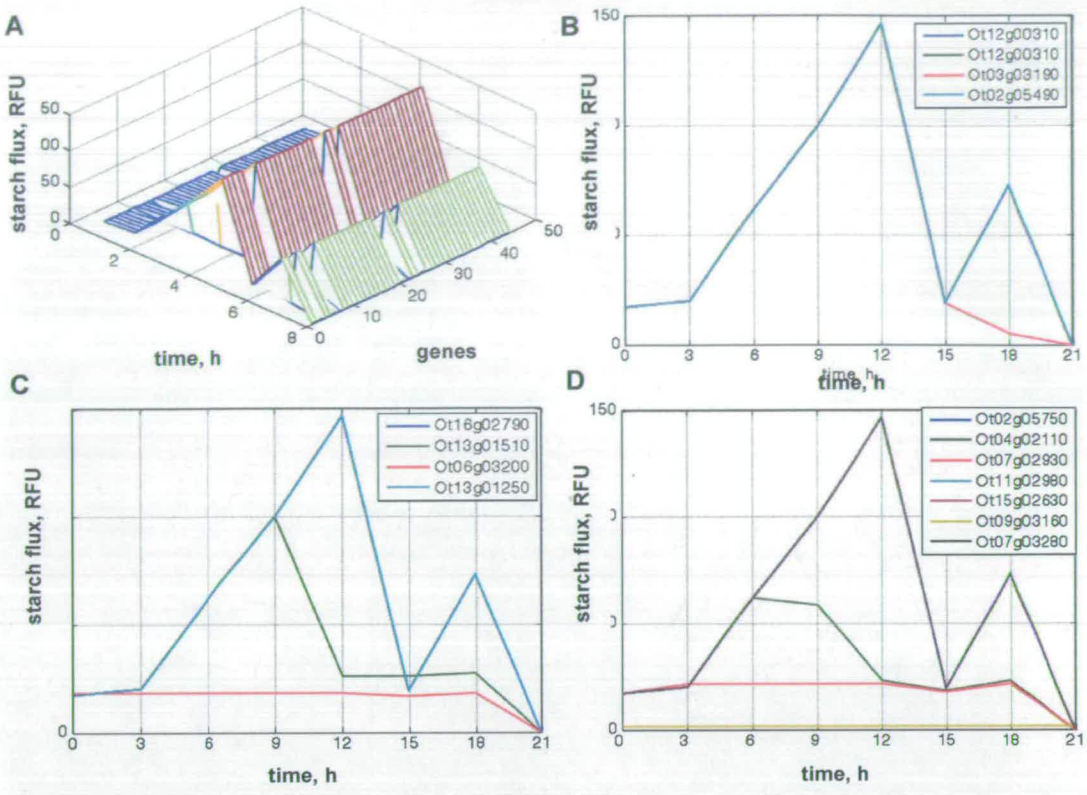


Figure 4.17 The effect of the *in silico* single gene deletion on the pattern of diurnal maltose production.

A. The 3D representation of the effect of single gene deletion on the shape of the maltose diurnal pattern.
B. The effect of the deletion of the Ot03g03190 (β- amylase) on the maltose diurnal pattern. Ot12g00310 (two profiles), Ot02g05490 correspond to the unperturbed shape and overlay each other.
C. The effect of the deletion of the Ot06g03200 (GBSSI) and Ot13g01510 (glucan, water dikinase) on the shape of the maltose diurnal pattern. Ot13g01250 and Ot16g02790 correspond to the unperturbed shape and overlay each other.
D. The effect of the deletion of the Ot07g02930 (AGPase), Ot0402110 (starch phosphorylase), Ot09g03160 (MEX1) on the shape of the maltose diurnal pattern. Ot07g03280, Ot02g05750, Ot15g02630 and Ot11g02980 correspond to the unperturbed shape and overlay each other.

are strongly supported by the well-known examples of plant starchless mutants [172-174]

Meanwhile, the deletion of the maltose transporter Ot09g03160 increases the rate of starch production. Interestingly, Ot04g02110, starch phosporylase deletion changes the pattern for the light-dark transition causing the significant

decrease in starch degradation rate during the dark period. From the last 5 genes we can observe Ot03g00330, fructose-1,6-bisphosphatase, decreasing the starch synthetic flux (Fig 4.16, D). This stems from the model structure where 1,6 bisphosphatase in the initial linear region of the pathway determines the rate of the glucose entrance for starch synthesis. In similar way, Ot10g01490, fructose-bisphosphate aldolase deletion blocks the starch synthesis during the day as it prevents the CO_2 input to the pathway.

For maltose the results of simulation of single gene deletion are presented at Figure 4.17. Here, the deletion of Ot03g03190 (β-amylase) abolishes the second maltose peak that happens at ZT18 (Fig 4.17, B). As the β-amylase is the main maltose supplier in the system, this effect is quite obvious. Both the deletions of Ot06g03200 (GBSSI) (Fig 4.17, C) and Ot07g02930 (AGPase) (Fig 4.17, D) remove the maltose pattern completely. These enzymes are the key payers in the starch production pathway, and their disappearance leads to significant decrease in the amount of the maltose precursor. Nevertheless, some nonzero maltose level remains due to maltose production from short glucans. Meanwhile, the deletion of the maltose transporter (MEX1) switches the maltose export off, for it blocks the possibility for maltose flux to occur.

Glucan, water dikinase Ot13g01510 defines the kinetics of 'dark' maltose region (Fig 4.17, C). This enzyme accounts for the starch granule phosphorylation, which precedes the starch degradation, thus, the deletion of the corresponding gene prevents the normal sequence of the events related to the starch degradation during the night. In the model, starch phosphorylase together with β–amylase constitute the two main sinks of the system; respectively, the availability of starch phosphorylase gene, Ot0402110, determines the magnitude of the first maltose peak and the existence of the second one (Fig 4.17, D), thus provides stronger effect that one might expect from the literature.

4.5.5. Single overexpression experiment

For overexpression each gene upper bound, one by one, was set to 2, and then, as in the general procedure, the combined effect of all genes' expression patterns has been estimated.

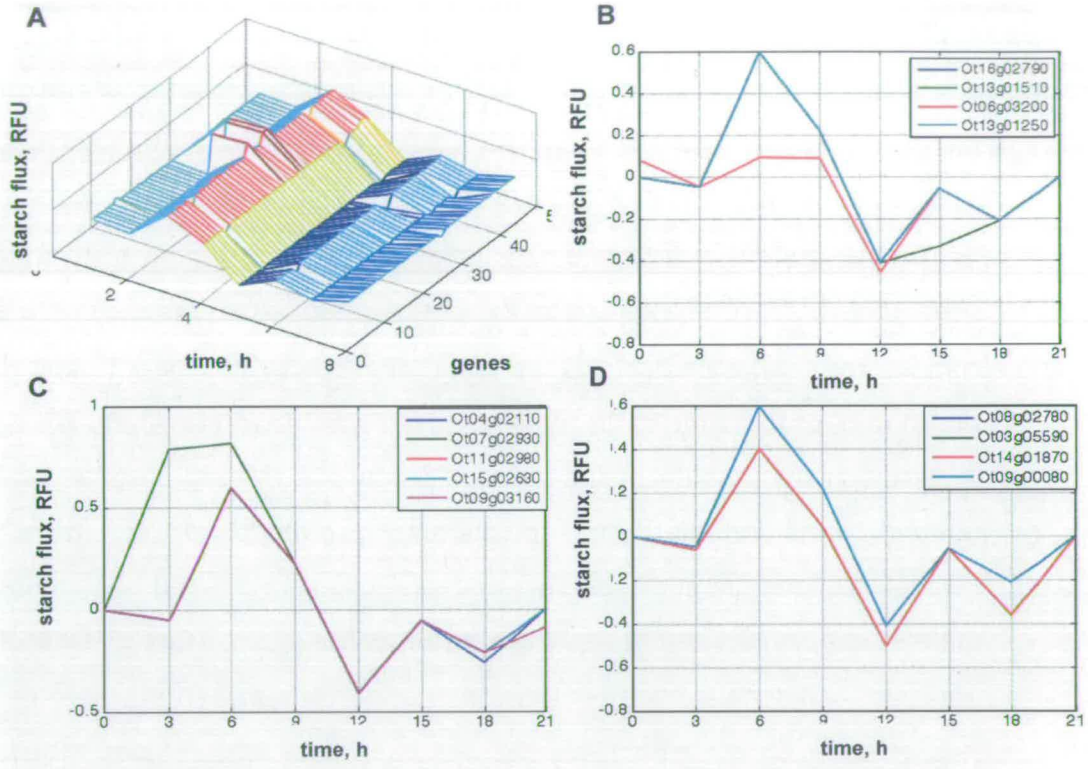


Figure 4.18 The effect of single gene overexpression on the diurnal pattern of the starch production.

A. The 3D representation of the results of the single gene overexpression experiment.

B. The effects of Ot06g03200 (GBSSI), Ot13g01510 (glucan, water dikinase) on the shape of diurnal starch production. Ot13g01250 and Ot16g02790 correspond to the unperturbed shape and overlay each other.

C. The overexpression of Ot07g02930 (AGPase) upregulates the starch production during the light period, while overexpression of Ot04g02110 (starch phosphorylase) and Ot09g03160 (maltose transporter) slightly affect the last two points of the night pattern. Ot15g02630, Ot11g02980 correspond to the unperturbed shape and overlay each other.

D. The effect of the overexpression of glucose transporters (Ot14g01870 and Ot03g05590, overlayed) on the overall diurnal pattern of the starch production. Ot09g00080 and Ot08g02780 correspond to the unperturbed shape and overlay each other.

As in previous simulations, the 3D picture gives the overall pattern of which genes' overexpression results in deviation of the starch pattern from the basic shape (Fig 4.18, A).

Surprisingly, GBSSI overexpression causes the result similar to its deletion: the decrease in starch flux during the subjective day. As it will be seen from the maltose simulation this coincides with the increasing flux for maltose production. In our model GBSSI is responsible for the elongation of long linear glucans. At the same time, we have a fixed ratio of linear to branched glucans in the content of our model starch. Thus, increasing flux for reactions carried out by Granule Bounded Starch Synthase (GBSSI) provides more long linear glucans than could be integrated to the starch granule and, respectively, they immediately become the available source for starch breakdown enzymes. All the glucan water dikinases (Ot13g01510, Ot08g01260 and Ot13g1250) straighten the 'dark' part of the pattern (Fig 4.18, B). This most likely happens because of enhanced starch phosphorylation and subsequent degradation in the dark.

Overexpression of AGPases (Ot07g02930 and Ot20g00490) predictably enhances the starch synthesis rate during the day (Fig 4.18, C). Upregulation of the glucose transporters (Ot14g01870 and Ot03g05590) gives the uniformly distributed lowering of starch flux level (Fig 4.18, D) due to the increased rate of the starch degradation and release from the chloroplast.

For maltose results of the single overexpression experiment are presented on the Figure 4.19.

The overexpression of Ot16g02370, Ot0g04170, Ot08g01260 and Ot13g01510 (glucan, water dikinases) change the dark pattern for maltose, resulting in a smoother decrease towards ZT21. Ot11g00280 and Ot11g01020 (starch phosphorylases) increase the magnitude of the second peak in the maltose flux pattern (Fig 4.19, B and C). Interestingly, for all the six genes mentioned here, we do not have the microarray information and their upper bound was set in the model at the minimum possible positive value: 0.01. The overexpression and, accordingly, the more pronounced participation of these genes in the

corresponding reactions change the underlying maltose diurnal flux. This gives grounds to suppose slightly different profiles for starch and maltose would be predicted in case of availability of microarray data, although the general trends will remain valid.

Overexpression of GBSSI increases the maltose flux during the day (Fig 4.19, C) and, as mentioned above, it corresponds to decrease in starch production rate.

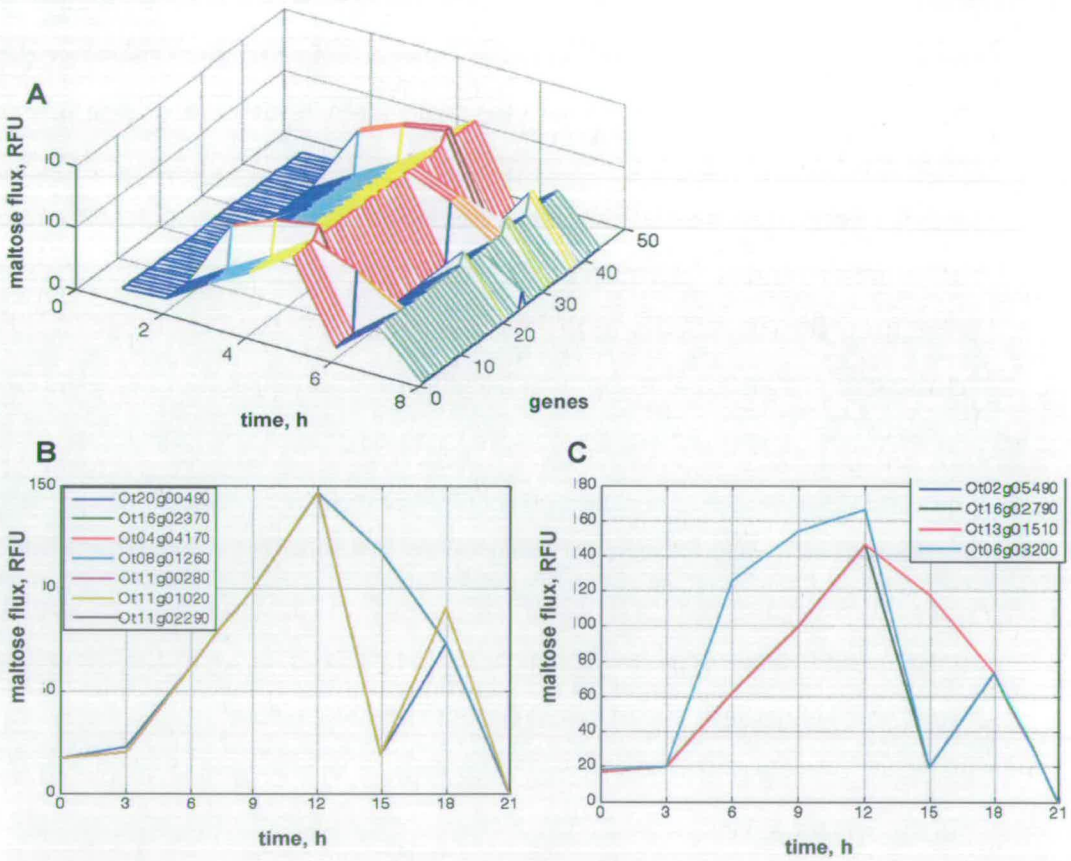


Figure 4.19 The effect of single gene overexpression on the diurnal pattern of the maltose production.

A. The 3D representation of the results of the single gene overexpression experiment.

B. The effects of the single overexpression of the Ot16g02370, Ot04g04170, Ot08g01260 (glucan, water dikinases, overlayed, light blew) and and Ot11g01020 (starch phosphorylases) on the diurnal shape of maltose production. Ot20g00490, Ot11g00280, Ot11g02290 represent the unperturbed shape and overlay each other.

C. The effects of single overexpression of the Ot13g01510 (glucan, water dikinase) and Ot06g03200 (GBSSI) on the diurnal shape of maltose production.Ot16g02790, Ot02g05490 correspond to unperturbed shape and overlay each other.

4.6. Discussion

The model structure and the constraints applied are sensible and are biologically plausible for they are based on the real pathway structure and described the main features of the starch metabolism.

The starch "phenotype" on a whole matches the data from [99] that starch partitioning and degradation occurs at nightfall, when the cell has accumulated sufficient biomass and polysaccharide storage for successful division. We have the increase in maltose outflow starting from ZT3, which has a peak at ZT12. Although we have no idea about the timing of actual enzyme activity, we can expect according to the starch and maltose profiles that there is no pronounced time delay between RNA and protein activity for the genes considered. This assumption is also supported by the experimental curve for the diurnal dynamics of starch content, which behaves in good agreement with model predictions.

Overall starch flux in our model is positive and equal to 0.4 FRU. This means that starch never degrades completely and starch production is always greater than degradation. This result is indirectly supported by the fact that nobody has observed the starch fully degraded in *Ostreococcus* and *Chlamydomonas* even within quite long periods of dark. However, I have also to stress that all the model predictions were calculated for the single cell, so the positive trends for the starch flux and content may arise from the fact that the dilution coefficient term has not been included in the model.

As a whole, the model has a predictive power in relation to the overall starch content estimating a peak approximately at ZT10, which matches the experimental data with a good accuracy level.

We revealed the set of potential targets of circadian regulation (Table F.1). The main player from the list: AGPase has already been reported as a key regulatory step for *Arabidopsis* and *Chlamydomonas*. The list also includes GBSSI, β-amylases, glucan water dikinases GWD (R1) and maltose transporter MEX1.

Interestingly, these targets have generally the same location relative to the pathway landscape as those identified to be redox regulated in plants (Fig 4.20).

As we observed, the main regulation targets are concentrated at the linear entrance to the system, within the actual topological bottleneck of the system (Fructose-1,6-bisphosphotase, ADPase, etc). Some of the events of high importance, corresponding to the synthesis and degradation, are located to granule surface. And here GBSSI and GWD appear in the foreground as the main potential targets for genetic regulation. The final set of targets deals with the production of maltose and its withdrawal from the system. The main target enzymes identified for this location are β–amylases and the maltose transporter MEX1.

The results of FBA robustness analysis are generally in agreement with the results of gene perturbations as the revealed genes successfully match the 'sensitive 'reactions. For instance, the GBSSI enzyme that appears in all the lists carries out reaction of 20LG glucan synthesis. Genes for debranching enzyme accompany the presence of debranching reactions like '100BGs_Pto20LG' and '60BGto20LG'. However, although the genes for glucan, water dikinase perform as attractive regulatory targets, the starch flux is not very sensitive to small fluctuations in fluxes of the corresponding reactions until they attain some significant threshold (Fig E.2: 36, 37, 38). Hence, for presumptive conclusions about the nature of regulation, both the genetic and also the biochemical and metabolomic data have to be taken into account.

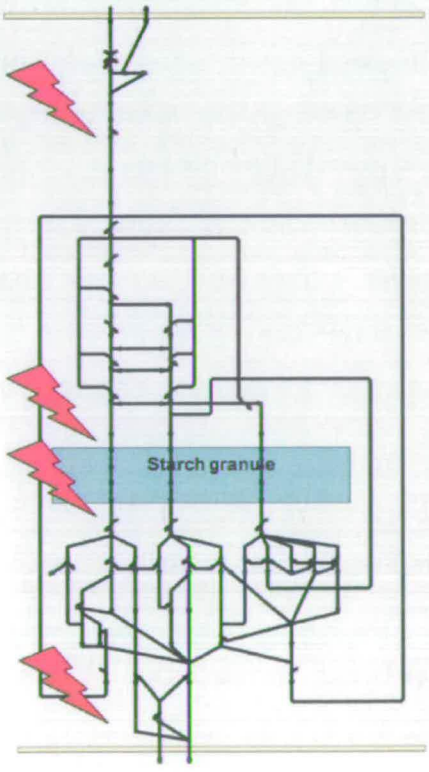


Figure 4.20 The location of the main target for genetic (circadian) regulation identified by simulations of single gene deletion and single gene overexpression.

The modelling approach presented here is hardly applicable to any plant system due to strong differences in the degree of posttranslational regulation and redox regulation in particular. Also, the mechanism of starch partitioning and division coinciding with the plastid division, as well as the ratio between the chloroplast volume and the volume of the cell seem to be the unique attributes of unicellular green algae. This, nevertheless, means that the method is applicable to systems of similar complexity like *Micromonas* or *Chlamydomonas.*

Conclusions

The work presented here demonstrates the advantages of the integrative multidisciplinary approach generally used by Systems Biology. Taking into account not only the properties of simple atomic units but all the spectrum of interactions between them opens additional prospects for the understanding of complex system behaviour. It allows a vision of the system not from the inside, which is typical in most experimental biology disciplines, but rather from outside, viewing the entire picture. The latter gives one a possibility to analyse the system's emergent properties and reveal the regularities and dependencies in the patterns observed.

In collaboration with experimentalists from the Szeged and Banyuls labs, I have achieved the construction of several models that describe all three considered components of circadian clock system: the light input, core oscillator and metabolic output pathways. In my work I have focused on mechanistic modelling, which takes into account prior knowledge about the mechanisms underlying the system's functioning and could be classified as deductive. For the first two subsystems I considered quantitative, mainly deterministic, kinetic modelling the most suitable as it gives the best fit for dynamic systems with a small number of components. In the last subsystem, the starch metabolic pathway, though also undoubtedly dynamic, demonstrates much more magnificent complexity; hence, I preferred the quasi- steady- state approach as most appropriate for my purposes. The implications of each model's design are discussed at the end of the corresponding chapter, but here I will summarize the key points of the work from the wider point of view of the Systems Biology approach.

The model for the PhyA- FHL light inducible system, actually, corresponds to the virtually ideal Systems Biology case, where the modelling and experimental work coexist in the same time scale and are mutually beneficial. By the time of the real implementation of the system in the yeast host, the preliminary model

based on Shimizu-Sato's data for the PhyB-PIF3 system [66] had already been developed. It predicted the most probable shape of the system's response to light induction. As soon as the first experimental curves for this model had been obtained, the model was fitted to them, whereupon new questions about the considered mechanism immediately arose. This is a quite common problem for Systems Biology: our available knowledge about the molecular mechanisms underlying particular physiological phenomena is quite incomplete. Here explanatory modelling appears to be really effective as it generates hypotheses that could be quickly examined for their reliability, and this at times helps to fill the gaps. The two most illustrative cases for the light inducible model from this point of view are the incorporation of the diffusion term in the model and the suggestion of the intermediate and moderately stable physical association between the inactive (Pr) form of PhyA and its interacting partner FHL. The first assumption is quite straightforward and stems from solid culture conditions. The introduction of the diffusion component to the model was essential to fit the data. The experiment that was designed according to the model hypothesis is presented in Fig 2.4. It confirmed the substantial influence of the diffusion process on the active concentration of luciferin in the solid yeast culture conditions and the necessity to take into account its influence on the overall performance of the system.

For the intermediate state Pr_FHL, its existence has been assumed exclusively on the basis of non- instantaneous kinetics of switching the system off. At the moment of the model development the evidence for signalling activity of the Pr form of PhyA was scant. Nevertheless, in recent publications Chen *et al*, 2009 and Yang *et al*, 2009 have reported that *in vivo* both FHL and FHY1 proteins interact preferentially with the Pr form of PhyA, although they are still able to interact with the Pfr form [175, 176]. This distinguishes PhyA-related signalling from PhyB, where the activity in the signalling cascade is attributed exclusively to the Pfr form. Moreover, more stable interaction between PhyA in Pr form and FHL/FHY1 was observed after subsequent 10-minute pulses of R/FR light. These data all support our hypothesis about the details of the light signalling mechanism in our system. Consequently, the choice of the PhyA-FHL interacting

pair for our system results in a light-attenuated rather than a light-switchable system, as could be seen from the simulation and experimental results.

The system, nevertheless, is able to provide a signal of variable length and amplitude to perturb any target promoter complex inside the cell. All parameters that define the characteristics of this signal are identified within the model. Therefore, the mode of the signal becomes a matter of directed design and could be estimated at will using the proposed model. This kind of light responsive module would be especially useful for the purposes of systematically perturbing an oscillatory system inside the cell, as in the case of our hypothetical synthetic core clock components.

From the point of view of Systems Biology, oscillations are a highly attractive and at the same time quite an informative phenomenon. On the one hand, sustained oscillations can be more easily detected in an experiment than, for instance, a small and rapid change in a steady state. The oscillation could also easily be broken and different perturbations could be interrogated with respect to the destruction and restoration of the oscillatory behaviour. On the other hand, in mathematics, the very strict and unambiguous bifurcation theory allows the formulation of the necessary conditions for giving rise to oscillations within the system. Indeed, oscillations are the product of the collective contribution of all the components of the network. Thus, reductionism is not appropriate as an approach for the study of the oscillatory systems. An examination of single atomic units of the system in isolation will never provide enough information for the detection of oscillations. Successful modelling of the natural circadian clock system so far reflects the advantage of the systems approach as it has already shed light on different aspects of molecular circadian clocks.

However, given the structural variability and overall complexity of the clock systems identified in mammals, flies and plants it seems attractive to build a minimal system that will retain the basic dynamic properties of the complicated ones. Recent successful applications of synthetic biology provide a new promising approach to complex network behaviour [60, 68, 69].. In Synthetic

Biology, networks are typically assembled from unrelated elements, which are accompanied by altering the kinetics of individual elements until they are impedance-matched such they function correctly within the new context. Here again modelling is becoming a powerful tool, allowing the testing of the structural topology and trying the different candidates for components without the labour of experimental work. As I have mentioned in the Introduction to Synthetic Biology, preliminary modelling appears as important as *in vitro* analysis of enzymatic activity in biochemistry. It allows experimentalists to concentrate on the most appropriate design and suggests ways for improving its performance.

In our work we examined the two potential designs for the core oscillator and chose one that demonstrates the desired behaviour in the biologically plausible parameter ranges. For this we, for the first time, successfully implemented the combination of the forward and inverse bifurcation analysis. Iterative use of forward and inverse bifurcation seems to be particularly profitable for networks with no prior knowledge about their dynamic behaviour, but with some available kinetic data reported for the individual components. Taking this kinetic information as a starting point in multidimensional parameter space, one may easily, in one operation, find the closest Hopf bifurcation by applying the Inverse Eigenvalue Analysis. From that point one could then move in the desired direction, following the parameter of interest with a Forward Bifurcation Analysis. In such a way, it is possible in a short time to achieve the most attractive (or biologically plausible) point in parameter space. Bifurcation Analysis also gives an idea of the relationship between the parameters in the model and provides information about the correlated parameters. For further experimental design this information is quite important as it allows prediction and avoidance of undesirable influences between the parameters; it also often provides experimentalists with a number of options for modification of the network design to that required for the desired behaviour.

After the synthetic system is designed and modelled it is ready for *in vivo* implementation. However, nobody would expect that, newly constructed in the

host cell, the synthetic network will perform as was expected by the developers. None of the existing synthetic systems to my knowledge demonstrated the exact desired mode of behaviour from the very beginning. Nevertheless, the results of experimental validation of the proposed design provide us with parameter values that will be incorporated into the model. It allows the further exploration of the system capacity and suggests further options for the experimental design. Using the Forward and Inverse Bifurcation analysis, we have intensively studied the parameter space for the Tet-On based model. In our case the most sensitive parameter revealed in both analyses is g_Ssn6, which corresponds to the Michaelis constant for the reaction of the transcription of the inhibitor component. This gives grounds to assume that if our system does not oscillate *in vivo*, the first point to focus on will be the binding affinity of the inhibitor component SSN6 for the promoter. Fortunately, mutating the DNA sequence of the cognate cis-element may readily control this. Further work on this project by the Szeged group will show how close to reality our predictions are.

We investigated the theoretical possibility for an oscillator of the proposed design to be affected with applied external force that had the same characteristics as our light-input system. The simulation revealed that the synthetic core oscillator could be easily entrained to a period close to its internal one. For our 12.5h oscillator the allowed range of entrainment is from 10 to 15 hours. We also tested how the combined contribution of various amplitudes and periods of the external signal affect the length of the resulting period. This potentially will provide some insight in the mechanisms of the light perception and daily synchronization of the natural circadian clock. Indeed, the synthetic and natural systems have significant differences e.g synthetic system has no rhythmic gating of light input that is common in evolved circadian clock, but that could also be engineered for future research. Knowing the exact profile of gene activation from the light switch it is easy to calculate the predicted phase-response curve from a set of single pulse experiments. Such an experiment will provide more information about the parameters of the oscillatory system than will brute-force fitting of the time series alone, as this

phase is more sensitive than the time series in terms of the parameter values [177, 178].

We have chosen the starch pathway for modelling of the clock output for several reasons. First of all, starch accumulation is an important mechanism for buffering of carbon fluctuations. Sucrose that has remobilized from starch during the night prevents carbon starvation in the absence of photosynthesis. Semicrystalline granule formation of starch is also assumed to be a protecting mechanism against the osmotic shock that may happen due to an excess of glucose emerging from carbon fixation during the light period of the day. This all means that the starch metabolism must be tightly clock regulated, as it strongly needs the precise linking to the time of light/dark transitions. Genetic regulation provides the availability of enough enzyme amount over the day, but the activity of these enzymes is a matter of post-translational regulation.

We expected that in *Ostreococcus* the degree of post-translational regulation would be lower, primarily because of the co-location of the storage and production processes. A highly compartmentalized plant has to cope with various demands of different cell, tissues and organs, which might be one of the reasons to evolve rapid responses to external metabolic feedbacks. Meanwhile, the unicellular alga might be able to survive with a less complicated mechanism of regulation. Redox regulation is capable of significantly moving the point of the metabolic steady state even in the presence of constant genetic input. As *O.tauri* lacks this regulation at least in the starch pathway, it becomes an ideal model for inferring circadian regulation from the time series transcriptome data. The modelling results obtained match the experimental data surprisingly well. Besides the validation of the model, this to some extent provides information about the properties of our clock "pendulum". To follow the core circadian oscillator precisely it should be impedance-balanced: in other words should be able to respond to rhythmic input from the clock without much delay and, hence, it shouldn't be over-burdened by complex post translational regulation similar to that observed in plants.

To conclude, we investigated different aspects of the circadian clock by considering minimal model systems. The effectiveness of this approach was demonstrated by the example of the model for the core synthetic oscillator, which provides the information about the general topology and parameter ranges that satisfy the desired dynamics and clearly could give insight into the mechanism of core clock-light input relationships. The minimal system for starch metabolism allowed us to gain some idea about the regulation of a key metabolic pathway and has revealed the potential topology of targets for circadian regulation, which show amazing analogy with that for redox–regulation in plants: the same enzymes that are redox-regulated in plants carry the limiting gene regulation in algae.. Nevertheless, one should be cautious directly approximating complicated systems from the minimal ones. The simpler systems can also have fundamental differences of design, such as the evident link of *Ostreococcus* starch metabolism to the cell division cycle.

Bibliography

1. Coveney PV, Fowler PW: **Modelling biological complexity: a physical scientist's perspective.** *J R Soc Interface* 2005, **2**:267-280
2. Jaqaman K, Danuser G: **Linking data to models: data regression.** *Nature Reviews Molecular Cell Biology* 2006, **7**:813-819.
3. Jong Hd: **Modeling and Simulation of Genetic Regulatory Systems: A Literature Review.** *Journal of Computational Biology* 2002, **9**:67-103. .
4. Palsson BO: **Systems Biology: Properties of Reconstructed Networks** *Cambridge Univ Press* 2006.
5. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK: **Physicochemical modelling of cell signalling pathways.** *Nature cell biology* 2006, **8**:1195-1203
6. Dunlap JC: **The Neurospora Circadian System** *Journal of Biological Rhythms* 2004, **19:** 414-424
7. Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, Nagy F, Hibberd JM, Millar AJ, Webb AAR: **Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage** *Science* 2005, **309**:630 - 633.
8. Eckardt NA: **A Wheel within a Wheel: Temperature Compensation of the Circadian Clock** *The Plant Cell* 2006, **18**:1105-1108.
9. Dunlap J, Loros J, Liu Y, Crosthwaite S: **Eukaryotic circadian systems: cycles in common** *GENES CELLS* 1999, **4**:1-10.
10. Schäfer E, Bowler C: **Phytochrome-mediated photoperception and signal transduction in higher plants.** *EMBO reports* 2002, **3**:1042–1048.
11. Staiger D, Allenbach L, Salathia N, Fiechter V, Davis SJ, Millar AJ, Chory J, Fankhauser C: **The Arabidopsis SRR1 gene mediates phyB signaling and is required for normal circadian clock function** *Genes & Dev* 2003, **17**:256-268
12. Sancar A: **Regulation of the Mammalian Circadian Clock by Cryptochrome.** *Journal of Biological Chemistry* 2004, **279**:4079-34082.
13. Collins BH, Rosato E, Kyriacou CP: **Seasonal behavior in Drosophila melanogaster requires the photoreceptors, the circadian clock, and phospholipase C.** *PNAS* 2004, **101**:1947-1950.
14. Gardner MJ, Hubbard KE, Hotta CT, Dodd AN, Webb AAR: **How plants tell the time.** *Biochem J* 2006, **397**:15–24.
15. Goodwin B: **Oscillatory behavior in enzymatic control processes.** *Adv Enzyme Regul* 1965, **3**:425-438.

16. Smolen P, Baxter DA, Byrne JH: **A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator.** *Biophys J* 2002, **83:**2349–2359.

17. Tsai TY-C, Choi YS, Ma W, Pomerening JR, Tang C, James E. Ferrell J: **Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops.** *Science* 2008, **321:**126 - 129.

18. Mehra A, Baker CL, Lorus JJ, Dunlap JC: **Post-tranlational modifications in circadian rhythms.** *Trends in Biochemical Sciences* 2009, **34:**483-490.

19. Young MW, Kay SA: **TIME ZONES: A COMPARATIVE GENETICS OF CIRCADIAN CLOCKS.** *Nature Reviews Genetics* 2001, **2:**702-715.

20. Fan J-Y, Preuss F, Muskus MJ, Bjes ES, Price JL: **Drosophila and Vertebrate Casein Kinase Iδ Exhibits Evolutionary Conservation of Circadian Function.** *Genetics* 2009, **181:**139–152.

21. Allada R, Meissner R-A: **Casein kinase 2, circadian clocks, and the flight from mutagenic light** *Molecular and Cellular Biochemistry* 2005, **274:**141-149.

22. Harmer SL, Hogenesch JB, Straume M, Chang H-S, Han B, Zhu T, Wang X, Kreps JA, Kay SA: **Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock** *Science* 2000, **290:**2110 - 2113.

23. Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL: **Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development.** *Genome Biology* 2008, **9:**R130.

24. Zhang EE, Liu AC, Hirota T, Miraglia LJ, Welch G, Pongsawakul PY, Liu X, Atwood A, III JWH, Janes J, et al: **A Genome-wide RNAi Screen for Modifiers of the Circadian Clock in Human Cells** *Cell* 2009, **139:**199-210

25. Lin Y, Stormo GD, Taghert PH: **The Neuropeptide Pigment-Dispersing Factor Coordinates Pacemaker Interactions in the Drosophila Circadian System** *The Journal of Neuroscience* 2004, **24:**7951-7957.

26. Dunlap JC, Loros JJ, DeCoursey PJ: **Chronobiology: Biological Timekeeping.** 2004.

27. Lee K, Loros JJ, Dunlap JC: **Interconnected Feedback Loops in the Neurospora Circadian System** *Science* 2000, **289** 107-110

28. Hong CI, Peter Ruoff, Loros JJ, Dunlap JC: **Closing the circadian negative feedback loop: FRQ-dependent clearance of WC-1 from the nucleus.** *Genes Dev* 2008, **22:**3196–3204. .

29. Glossop NRJ, Lyons LC, Hardin PE: **Interlocked Feedback Loops Within the Drosophila Circadian Oscillator** *Science* 1999, **286:**766 - 768.

30. Lee C, Bae K, Edery I: **PER and TIM Inhibit the DNA Binding Activity of a Drosophila CLOCK-CYC/dBMAL1 Heterodimer without Disrupting Formation of the Heterodimer: a Basis for Circadian Transcription** *Molecular and Cellular Biology* 1999, **19:**5316-5325.

31. Kim EY, Edery I: **Balance between DBT/CKIε kinase and protein phosphatase activities regulate phosphorylation and stability of Drosophila CLOCK protein** *PNAS* 2006, **103:**6178-6183

32. Goldbeter A: **A Model for Circadian Oscillations in the Drosophila Period Protein (PER).** *Proceedings: Biological Sciences* 1995, **261:**319-324

33. Cyran SA, Buchsbaum AM, Reddy KL, Lin M-C, Glossop NRJ, Hardin PE, Young MW, Storti RV, Blau J: **vrille, Pdp1, and dClock Form a Second Feedback Loop in the Drosophila Circadian Clock.** *Cell* 2003, **112:**329-341.

34. Benito J, Hao Zheng, Hardin PE: **PDP1{varepsilon} Functions Downstream of the Circadian Oscillator to Mediate Behavioral Rhythms.** *The Journal of Neuroscience* 2007, **27:**2539-2547.

35. Myers MP, Wager-Smith K, Rothenfluh-Hilfiker A, Young MW: **Light-Induced Degradation of TIMELESS and Entrainment of the Drosophila Circadian Clock** *Science* 1996, **271:**1736 - 1740.

36. Toh KL: **Basic Science Review on Circadian Rhythm Biology and Circadian Sleep Disorders.** *Ann Acad Med Singapore* 2008, **37::**662-667.

37. Guillaumond F, Dardente H, Giguère V, Cermakian N: **Differential Control of Bmal1 Circadian Transcription by REV-ERB and ROR Nuclear Receptors** *Journal of Biological Rhythms* 2005, **20:**391-403.

38. Bagyan S, Mair T, Dulos E, Boissonade J, Kepper PD, Müller SC: **Glycolytic oscillations and waves in an open spatial reactor: Impact of feedback regulation of phosphofructokinase** *Biophysical Chemistry* 2005, **116:**67-76

39. Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, Kondo T: **Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro.** *Science* 2005, **308:**414 - 415.

40. Murayama Y, Oyama T, Kondo T: **Regulation of Circadian Clock Gene Expression by Phosphorylation States of KaiC in Cyanobacteria.** *Journal of Bacteriology,* 2008, **190:**1691-1698.

41. McClung CR: **Plant Circadian Rhythms** *The Plant Cell* 2006, **18:**792-803.

42. Alabadí D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA: **Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock.** *Science* 2001, **293:** 880-883.

43. Daniel X, Sugano S, Tobin EM: **CK2 phosphorylation of CCA1 is necessary for its circadian oscillator function in Arabidopsis.** *PNAS* 2004, **101:**3292-3297.

44. Farré EM, Harmer SL, Harmon FG, Yanovsky MJ, Kay SA: **Overlapping and Distinct Roles of PRR7 and PRR9 in the Arabidopsis Circadian Clock.** *Current Biology* 2005, **15:**47-54.

45. Locke JCW, Kozma-Bognár L, Gould PD, Fehér B, Keve É, Nagy F, Turner MS, Hall A, Millar AJ: **Experimental validation of a predicted feedback**

loop in the multi-oscillator clock of Arabidopsis thaliana. *Mol Syst Biol* 2006, **2**:59.

46. Doyle MR, Davis SJ, Bastow RM, McWatters HG, Kozma-Bognár L, Nagy F, Millar AJ, Amasino RM: **The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana.** *Nature* 2002, **419**:74-77.
47. McWatters HG, Bastow RM, Hall A, Millar AJ: **The ELF3zeitnehmer regulates light signalling to the circadian clock.** *Nature* 2000, **408**:716-720.
48. Hazen SP, Schultz TF, Pruneda-Paz JL, Borevitz JO, Ecker JR, Kay SA: **LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms** *PNAS* 2005, **102**:10387-10392
49. Smolen P, Hardin PE, Lo BS, Baxter DA, Byrne JH: **Simulation of Drosophila Circadian Oscillations, Mutations, and Light Responses by a Model with VRI, PDP-1, and CLK** *Biophysical Journal* 2004, **86**:2786-2802
50. Leloup J-C, Goldbeter A: **Toward a detailed computational model for the mammalian circadian clock.** *PNAS* 2003, **100**:7051-7056
51. François P: **A Model for the Neurospora Circadian Clock.** *Biophysical Journal* 2005, **88**:2369-2383.
52. Locke JCW, Southern MM, Kozma-Bognár L, Hibberd V, Brown PE, Turner MS, Millar AJ: **Extension of a genetic network model by iterative experimentation and mathematical analysis.** *Molecular Systems Biology* 2005, **1**.
53. UEDA HR, HAGIWARA M, KITANO H: **Robust Oscillations within the Interlocked Feedback Model of Drosophila Circadian Rhythm** *Journal of Theoretical Biology* 2001, **210**:401-406
54. Ruoff P, Christensen MK, Sharma VK: **PER/TIM-mediated amplification, gene dosage effects and temperature compensation in an interlocking-feedback loop model of the Drosophila circadian clock** *Journal of Theoretical Biology* 2005, **237**:41-57
55. Forger DB, Peskin CS: **A detailed predictive model of the mammalian circadian clock.** *PNAS* 2003, **100**:14806-14811.
56. Forger DB, Peskin CS: **Stochastic simulation of the mammalian circadian clock.** *PNAS* 2005, **102**:321-324.
57. Heinemann M, Panke S: **Synthetic biology—putting engineering into biology** *Bioinformatics* 2006, **22**:2790-2799.
58. Kaznessis YN: **Models for synthetic biology.** *BMC Syst Biol* 2007, **1**:47.
59. Andrianantoandro E, Basu S, Karig DK, Weiss R: **Synthetic biology: new engineering rules for an emerging discipline.** *Mol Syst Biol* 2006, **2**:2006.0028. .
60. Wong WW, Liao JC: **The design of intracellular oscillators that interact with metabolism.** *Cellular and molecular life sciences:CMLS* 2006, **63**:1215-1220.

61. Kærn M, Blake WJ, Collins JJ: **THE ENGINEERING OF GENE REGULATORY NETWORKS.** *Annu Rev Biomed Eng* 2003, **5**:179–206.

62. Bellí G, Garí E, Piedrafita L, Aldea M, Herrero E: **An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast.** *Nucleic Acids Res* 1998, **26**:942–947.

63. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature* 2000, **403**:335-338

64. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli.** *Nature* 2000, **403**:339-342

65. Gossen M, Freundlieb S, Bender G, Müller G, Hillen W, Bujard H: **Transcriptional activation by tetracyclines in mammalian cells.** *Science* 1995, **268**:1766-1769.

66. Shimizu-Sato S, Huq E, Tepperman JM, Quail PH: **A light-switchable gene promoter system.** *Nature Biotechnology* 2002, **20**:1041 - 1044.

67. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M: **A tunable synthetic mammalian oscillator.** *Nature* 2009, **457**:309-312.

68. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ: **Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli** *Cell* 2003, **113**:597-607

69. Fung E, Wong WW, Suen JK, Bulter T, Lee S-g, Liao JC: **A synthetic gene–metabolic oscillator.** *Nature* 2005, **435**:118-122

70. Quail PH: **Phytochrome photosensory signalling networks.** *Nature Reviews Molecular Cell Biology* 2002, **3**:85-93.

71. Nagy F, Schäfer E: **PHYTOCHROMES CONTROL PHOTOMORPHOGENESIS BY DIFFERENTIALLY REGULATED, INTERACTING SIGNALING PATHWAYS IN HIGHER PLANTS.** *Annual Review of Plant Biology* 2002, **53**:329-355.

72. Hennig L, Büche C, Eichenberg K, Schäfer E: **Dynamic Properties of Endogenous Phytochrome A in Arabidopsis Seedlings.** *Plant Physiol* 1999, **121**:571–578.

73. Zhu Y, Tepperman JM, Fairchild CD, Quail PH: **Phytochrome B binds with greater apparent affinity than phytochrome A to the basic helix–loop–helix factor PIF3 in a reaction requiring the PAS domain of PIF3.** *PNAS* 2000, **97**:13419–13424. .

74. Al-Sady B, Ni W, Kircher S, Schäfer E, Quail PH: **Photoactivated Phytochrome Induces Rapid PIF3 Phosphorylation Prior to Proteasome-Mediated Degradation.** *Molecular Cell* 2006, **23**:439-446.

75. Ni M, Tepperman JM, Quai PH: **Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light.** *Nature* 1999, **400**:781-784.

76. Eichenberg K, Kunkel T, Kretsch T, Speth V, Schäfer E: **In Vivo Characterization of Chimeric Phytochromes in Yeast** *J Biol Chem* 1999, **274**:354-359.

77. Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M, Davidson EA, Scouras A, Ellington AD, Marcotte EM, Voigt CA: **Synthetic biology: Engineering Escherichia coli to see light.** *Nature* 2005, **438:**441-442.

78. Wang H, Deng XW: **Dissecting the phytochrome A-dependent signaling network in higher plants.** *TRENDS in Plant Science* 2003, **8:**172-178.

79. Leung DW, Otomo C, Chory J, Rosen MK: **Genetically encoded photoswitching of actin assembly through the Cdc42-WASP-Arp2/3 complex pathway.** *PNAS* 2008, **105:**12797-12802

80. Grennan AK: **Regulation of Starch Metabolism in Arabidopsis Leaves** *Plant Physiology* 2006, **142:**1343-1345

81. Ball SG, Morell MK: **FROM BACTERIAL GLYCOGEN TO STARCH: Understanding the Biogenesis of the Plant Starch Granule.** *Annual Review of Plant Biology* 2003, **54:**207-233.

82. Deschamps P, Moreau H, Worden{ddagger AZ, Dauvillée D, Ball SG: **Early Gene Duplication Within Chloroplastida and Its Correspondence With Relocation of Starch Metabolism to Chloroplasts.** *Genetics* 2008, **178:**2373-2387.

83. Usadel B, Bläsing OE, Gibon Y, Retzlaff K, Höhne M, Günther M, Stitt M: **Global Transcript Levels Respond to Small Changes of the Carbon Status during Progressive Exhaustion of Carbohydrates in Arabidopsis Rosettes1,[W],[OA].** *Plant Physiology* 2008, **146:**1834-1861.

84. Smith AM, Stitt M: **Coordination of carbon supply and plant growth.** *Plant, Cell and Environment* 2007, **30:**1126-1149.

85. Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, Hylton C, Zeeman SC, Smith AM: **Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in Arabidopsis Leaves.** *Plant Physiol* 2004, **136:**2687–2699.

86. Zeeman SC, Tiessen A, Pilling E, Kato KL, Donald AM, Smith AM: **Starch Synthesis in Arabidopsis. Granule Synthesis, Composition, and Structure.** *Plant Physiol* 2002, **129:**516–529.

87. Bläsing OE, Gibon Y, Günther M, Höhne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible W-R, Stitt M: **Sugars and Circadian Regulation Make Major Contributions to the Global Regulation of Diurnal Gene Expression in ArabidopsisW in Box.** *Plant Cell* 2005, **17:**3257–3281.

88. Lu Y, Gehan JP, Sharkey TD: **Daylength and Circadian Effects on Starch Degradation and Maltose Metabolism.** *Plant Physiol* 2005, **138:** 2280–2291.

89. Orzechowski S: **Starch metabolism in leaves.** *Acta Biochimica Polonica* 2008, **55:**435-445.

90. Geigenberger P, Kolbe A, Tiessen A: **Redox regulation of carbon storage and partitioning in response to light and sugars.** *Journal of Experimental Botany* 2005, **56:**1469-1479.

91. Tiessen A, Hendriks JHM, Stitt M, Branscheid A, Gibon Y, Farré EM, Geigenberger P: **Starch Synthesis in Potato Tubers Is Regulated by Post-Translational Redox Modification of ADP-Glucose Pyrophosphorylase.** *The Plant Cell* 2002, **14:** 2191-2213.

92. Mikkelsen R, Mutenda KE, Mant A, Schürmann P, Blennow A: **α-Glucan, water dikinase (GWD): A plastidic enzyme with redox-regulated and coordinated catalytic activity and binding affinity.** *PNAS* 2005, **102:**1785-1790

93. Sparla F, Costa A, Schiavo FL, Pupillo P, Trost P: **Redox Regulation of a Novel Plastid-Targeted β-Amylase of Arabidopsis.** *Plant Physiology* 2006, **141:**840–850.

94. Oesterhelt C, Klocke S, Holtgrefe S, Linke V, Weber APM, Scheibe R: **Redox Regulation of Chloroplast Enzymes in Galdieria sulphuraria in View of Eukaryotic Evolution.** *Plant and Cell Physiology* 2007, **48:**1359-1373;.

95. Peers G, Niyogi KK: **Pond Scum Genomics: The Genomes of Chlamydomonas and Ostreococcus.** *The Plant Cell* 2008, **20:**502-507.

96. Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynié S, Cooke R, et al: **Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features.** *PNAS* 2006, **103:**11647–11652.

97. Henderson GP, Gan L, Jensen GJ: **3-D Ultrastructure of O. tauri: Electron Cryotomography of an Entire Eukaryotic Cell.** *PLoS ONE* 2007, **8.**

98. Fouilland E, Descolas-Gros C, Courties C, Collos Y, Vaquer A, Gasc A: **Productivity and growth of a natural population of the smallest free-living eukaryote under nitrogen deficiency and sufficiency.** *Microbial ecology* 2004 **48:**103-110.

99. Ral J-P, Derelle E, Ferraz C, Wattebled F, Farinas B, Corellou F, Buléon A, Slomianny M-C, Delvalle D, d'Hulst C, et al: **Starch Division and Partitioning. A Mechanism for Granule Propagation and Maintenance in the Picophytoplanktonic Green Alga Ostreococcus tauri.** *Plant Physiol* 2004, **136:**3333–3340.

100. Ral J-P, Colleoni C, Wattebled F, Dauvillée D, Nempont C, Deschamps P, Li Z, Morell MK, Chibbar R, Purton S, et al: **Circadian Clock Regulation of Starch Metabolism Establishes GBSSI as a Major Contributor to Amylopectin Synthesis in Chlamydomonas reinhardtii.** *Plant Physiology* 2006, **142:**305-317.

101. Moulager M, Monnier A, Jesson B, Bouvet R, Mosser J, Schwartz C, Garnier L, Corellou F, Bouget F-Y: **Light-Dependent Regulation of Cell Division in Ostreococcus: Evidence for a Major Transcriptional Input.** *Plant Physiol* 2007, **144:**1360–1369.

102. Goto K, Johnson C: **Is the cell division cycle gated by a circadian clock? The case of Chlamydomonas reinhardtii.** *The Journal of Cell Biology,* 1995, **129:**1061-1069.

103. Françoisa J, Parroua JL: **Reserve carbohydrates metabolism in the yeast Saccharomyces cerevisiae.** *FEMS Microbiology Reviews* 2006, **25:**125 - 145.

104. Quail P, Boylan M, Parks B, Short T, Xu Y, Wagner D: **Phytochromes: photosensory perception and signal transduction** *Science* 1995, **268:**675-680.

105. Kim J, Yi H, Choi G, Shin B, Song P-S, Choi G: **Functional Characterization of Phytochrome Interacting Factor 3 in Phytochrome-Mediated Light Signal Transduction.** *The Plant Cell* 2003, **15:**2399–2407.

106. Hiltbrunner A, VICZIAN A, BURY E, TSCHEUSCHLER A, KIRCHER S, TOTH R, HONSBERGER A, NAGY F, FANKHAUSER C, SCHÄFER E: **Nuclear accumulation of the phytochrome A photoreceptor requires FHY1.** *Current biology* 2005, **15:**2125-2130

107. Hiltbrunner A, TSCHEUSCHLER A, VICZIAN A, KUNKEL T, KIRCHER S, SCHÄFER E: **FHY1 and FHL act together to mediate nuclear accumulation of the phytochrome A photoreceptor.** *Plant and cell physiology* 2006, **47:**1023-1034

108. Bachmair A, Finley D, Varshavsky A: **In vivo half-life of a protein is a function of its amino-terminal residue.** *Science* 1986, **234:**179-186.

109. McNabb DS, Reed R, Marciniak RA: **Dual Luciferase Assay System for Rapid Assessment of Gene Expression in Saccharomyces cerevisiae** *Eukaryotic Cell* 2005, **4:**1539-1549.

110. Robertson JB, Stowers CC, Boczko E, Johnson CH: **Real-time luminescence monitoring of cell-cycle and respiratory oscillations in yeast.** *PNAS* 2008, **105:**17988-17993

111. Novère NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, et al: **The Systems Biology Graphical Notation.** *Nature Biotechnology* 2009, **27:**735 - 741

112. Sorokina O, Kapus A, Terecskei K, Dixon LE, Kozma-Bognar L, Nagy F, Millar AJ: **A switchable light-input, light-output system modelled and constructed in yeast.** *J Biol Eng* 2009, **3:**15.

113. Li L, LAGARIAS JC: **Phytochrome assembly in living cells of the yeast Saccharomyces cerevisiae.** *Plant Biology* 1994, **91:**12535-12539.

114. Kunkel T, Speth V, Büche C, Schäfer E: **In Vivo Characterization of Phytochrome-Phycocyanobilin Adducts in Yeast** *Journal of Biological Chemistry* 1995, **270:**20193-20200.

115. Nagy F, Kircher S, Schafer E: **Intracellular trafficking of photoreceptors during light-induced signal transduction in plants.** *Journal of Cell Science* 2001 **114:**475-480.

116. Kendrick RE, Cronenberg GHM: **Photomorphogenesis in Plants.** 1994.

117. Hennig L, Schäfer E: **Both subunits of the dimeric plant photoreceptor phytochrome require chomophore for Pfr stability** *JBC Papers in Press*

2000.

118. Ruddat A, Schmidt P, Gatz C, Braslavsky SE, Gärtner W, Schaffner K: **Recombinant Type A and B Phytochromes from Potato. Transient Absorption Spectroscopy.** *Biochemistry* 1996, **36:**103-111.

119. Ignowski JM, Schaffer DV: **Kinetic analysis and modeling of firefly luciferase as a quantitative reporter gene in live mammalian cells.** *Biotechnology and Bioingeneering* 2004, **86:**827834.

120. Bonev B, Hooper J, Parisot J: **Principles of assessing bacterial susceptibility to antibiotics using the agar diffusion method.** *Journal of Antimicrobial Chemotherapy* 2008.

121. Arnikar H, Tattitali S: **Diffusion of rose bengal in agar gel: Variation with the concentration of the diffusant and the gel.** *Journal of Radioanalytical and Nuclear Chemistry* 1986, **105:**193-199.

122. McCabe M: **The diffusion coefficient of caffeine through agar gels containing a hyaluronic acid-protein complex. A model system for the study of the permeability of connective tissues.** *The Biochemical journal* 1972, **127:**249-253.

123. Westrin B: **Measurement of the diffusion coefficient of ethanol in agarose gel beads: A reproducibility study.** *Biotechnology Techniques* 1990, **4:**409-414.

124. THOMPSON JF, HAYES LS, LLOYD DB: **Modulation of firefly luciferase stability and impact on studies of gene regulation.** *Gene* 1991, **103:**171-177.

125. Finkenstädt B, Heron EA, Komorowski M, Edwards K, Tang S, Harper CV, Davis JRE, White MRH, Millar AJ, Rand DA: **Reconstruction of transcriptional dynamics from gene reporter data using differential equations.** *Bioinformatics* 2008, **24:**2901-2907.

126. Schmidt H, Jirstrand M: **Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology.** *BIOINFORMATICS* 2006, **22:**514-515.

127. Banks A, Vincent J, Anyakoha C: **A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications** *Natural Computing: an international journal* 2008, **7**.

128. Hengl S, Kreutz C, Timmer J, Maiwald T: **Data-based identifiability analysis of non-linear dynamical models.** *BIOINFORMATICS* 2007, **23:**2612–2618.

129. Cantone I, Marucci L, Iorio F, Ricci MA, Belcastro V, Bansal M, Santini S, Bernardo Md, Bernardo Dd, Cosma MP: **A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches** *Cell* 2009, **137:**172-181

130. Bhat PJ, Oh D, Hopper JE: **Analysis of the GAL3 Signal Transduction Pathway Activating GAL4 Protein-Dependent Transcription in Saccharomyces cerevisiae.** *Genetics* 1990, **125**:281-291.

131. Peng G, Hopper JE: **Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein.** *PNAS* 2002, **99**:8548-8553

132. Verma M, Bhat PJ, Venkatesh KV: **Quantitative Analysis of GAL Genetic Switch of Saccharomyces cerevisiae Reveals That Nucleocytoplasmic Shuttling of Gal80p Results in a Highly Sensitive Response to GalactoseQuantitative Analysis of GAL Genetic Switch of Saccharomyces cerevisiae Reveals That Nucleocytoplasmic Shuttling of Gal80p Results in a Highly Sensitive Response to Galactose.** *J Biol Chem* 2003, **278**:48764-48769.

133. http://web.wi.mit.edu/young/expression/halflife.html.

134. Melcher K, Xu HE: **Gal80–Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression.** *EMBO J* 2001, **20**:841–851.

135. Venkatesh KV, Bhat PJ, Kumar RA, Doshi P: **Quantitative Model for Gal4p-Mediated Expression of the Galactose/Melibiose Regulon in Saccharomyces cerevisiae.** *Biotechnology Progress* 1999, **15**:51 - 57.

136. Smidtas S, Schächter V, Képès F: **The adaptive filter of the yeast galactose pathway** *Journal of Theoretical Biology* 2006, **242**:372-381

137. Garí E, Piedrafita L, Aldea M, Herrero E: **A Set of Vectors with a Tetracycline-Regulatable Promoter System for Modulated Gene Expression in Saccharomyces cerevisiae.** *Yeast* 1998, **13**:837 - 848.

138. Davie JK, Trumbly RJ, Dent SYR: **Histone-Dependent Association of Tup1-Ssn6 with Repressed Genes In Vivo** *Molecular and Cellular Biology* 2002, **22**:693-703.

139. BARON U, BUJARD H: **Tet Repressor-Based System for Regulated Gene Expression in Eukaryotic Cells: Principles and Advances.** *METHODS IN ENZYMOLOGY* 2000, **327**:401-421.

140. Baron U, Gossen M, Bujard H: **Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential.** *Nucleic Acids Research* 1997, **25**:2723-2729.

141. Wishart JA, Hayes A, Wardleworth L, Zhang N, Oliver SG: **Doxycycline, the drug used to control the tet-regulatable promoter system, has no effect on global gene expression in Saccharomyces cerevisiae.** *Yeast* 2005, **22**:565 - 569.

142. Zhou X, Symons J, Hoppes R, Krueger C, Berens C, Hillen W, Berkhout B, Das AT: **Improved single-chain transactivators of the Tet-On gene expression system.** *BMC Biotechnol* 2007, **7**.

143. Peña MMO, Koch KA, Thiele DJ: **Dynamic Regulation of Copper Uptake and Detoxification Genes in Saccharomyces cerevisiae** *Mol Cell Biol* 1998, **18**:2514-2523.

144. Tzamarias D, Struhl K: **Functional dissection of the yeast Cyc8–Tupl transcriptional co-repressor complex.** *Nature* 1994, **369:**758 - 761

145. VARANASI US, KLIS M, MIKESELL PB, TRUMBLY RJ: **The Cyc8 (Ssn6)-Tup1 Corepressor Complex Is Composed of One Cyc8 and Four Tup1 Subunits.** *MOLECULAR AND CELLULAR BIOLOGY* 1006, **16:**6707–6714 Vol.

146. Keleher CA, Redd MJ, Schultz J, Carlson M, Johnson AD: **Ssn6-Tup1 is a general repressor of transcription in yeast** *Cell* 1992, **68:**709-719

147. Kedracka-Krok S, Gorecki A, Bonarek P, Wasylewski Z: **Kinetic and Thermodynamic Studies of Tet Repressor–Tetracycline Interaction.** *Biochemistry* 2005, **44:**1037–1046.

148. Takahashi M, Altschmied L, Hillen W: **Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements *1: Evidence for divalent metal ion requirement and energy transfer** *Journal of Molecular Biology* 1986, **187:**341-348

149. Kleinschmidt C, Tovar K, Hillen W, Porschke D: **Dynamics of repressor-operator recognition: Tn10-encoded tetracycline resistance control.** *Biochemistry* 1988, **27:**1094–1104.

150. Culotta VC, Hsu T, Hu S, Fürst P, Hamer D: **Copper and the ACE1 regulatory protein reversibly induce yeast metallothionein gene transcription in a mouse extract.** *PNAS* 1989, **86:**8377–8381.

151. Lin C, Kosman D: **Copper uptake in wild type and copper metallothionein-deficient Saccharomyces cerevisiae. Kinetics and mechanism.** *J Biol Chem* 1990, **265:**9194-9200.

152. http://oscill8.sourceforge.net/.

153. Conrad ED: **Bifurcation Analysis and Qualitative Optimization of Models in Molecular Cell Biology with Applications to the Circadian Clock.** *Ph D Thesis* 2006.

154. Lu J, Engl HW, Schuster P: **Inverse bifurcation analysis: application to simple gene systems.** *Algorithms Mol Biol* 2006, **1**.

155. Lu J: **Inverse Eigenvalue Problems for Exploring theDynamics of Systems Biology Models".** *Advances in Applied Mathmatics and Mechanics* 2009, **to appear**.

156. Winfree AT: **Integrated view of resetting a circadian clock.** *Journal of Theoretical Biology* 1970, **28:**327-374

157. Varma A, Palsson BO: **Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use.** *Bio/Technology* 1994, **12:**994 - 998.

158. Wang Q, Chen X, Yang Y, Zhao X: **Genome-scale in silico aided metabolic analysis and flux comparisons of Escherichia coli to improve succinate production.** *Applied Microbiology and Biotechnology* 2006, **73:**887-894.

159. Duarte NC, Herrgård MJ, Palsson BØ: **Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model.** *Genome Res* 2004, **14:**1298–1309..

160. Boyle NR, Morgan JA: **Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii.** *BMC Systems Biology* 2009, **3.**

161. Raman K, Rajagopalan P, Chandra N: **Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs.** *PLoS Comput Biol* 2005, **1(5).**

162. Schwender J, Ohlrogge JB, Shachar-Hill Y: **A Flux Model of Glycolysis and the Oxidative Pentosephosphate Pathway in Developing Brassica napus Embryos.** *J Biol Chem* 2003, **278:**29442-29453.

163. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E: **Network-based prediction of human tissue-specific metabolism.** *Nature Biotechnology* 2008, **26:**1003 - 1010.

164. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng T-Y, Moody DB, Murray M, Galagan JE: **Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production.** *PLoS Computational Biology* 2009, **5:**e1000489.

165. Zabawinski C, Koornhuyse NVD, D'Hulst C, Schlichting R, Giersch C, Delrue B, Lacroix J-M, Preiss J, Ball S: **Starchless Mutants of Chlamydomonas reinhardtii Lack the Small Subunit of a Heterotetrameric ADP-Glucose Pyrophosphorylase** *Journal of Bacteriology,* 2001, **183:**1069-1077.

166. Zeeman SC, Critchley JH, Takaha T, Smith SM, Smith AM: **The synthesis and degradation of starch in Arabidopsis leaves: the role of disproportionating enzyme.** *Starch: advances in structure and function Proceedings of Starch 2000: Structure and Function, Cambridge, UK, 27-29 March 2000* 2000.

167. Zhang X, Myers AM, James MG: **Mutations Affecting Starch Synthase III in Arabidopsis Alter Leaf Starch Structure and Increase the Rate of Starch Synthesis.** *Plant Physiology* 2005, **138:**663-674.

168. Dumez S, Wattebled F, Dauvillee D, Delvalle D, Planchotb V, Ball SG, D'Hulst C: **Mutants of Arabidopsis Lacking Starch Branching Enzyme II Substitute Plastidial Starch Synthesis by Cytoplasmic Maltose Accumulation.** *The Plant Cell* 2006, **18:**2694-2709.

169. Niittylä T, Messerli G, Trevisan M, Chen J, Smith AM, Zeeman SC: **A Previously Unknown Maltose Transporter Essential for Starch Degradation in Leaves.** *Science* 2004, **303:**87 - 89.

170. Weber A, Servaites JC, Geiger DR, Kofler H, Hille D, Gröner F, Hebbeker U, Flügge U-I: **Identification, Purification, and Molecular Cloning of a Putative Plastidic Glucose Translocator.** *Plant Cell* 2000, **12:**787–802.

171. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.** *Nature Protocols* 2007, **2:**727 - 738

172. Caspar T, Huber SC, Somerville C: **Alterations in Growth, Photosynthesis, and Respiration in a Starchless Mutant of Arabidopsis thaliana (L.) Deficient in Chloroplast Phosphoglucomutase Activity.** *Plant Physiology* 1985, **79:**11-17.

173. Lin T-P, Caspar T, Somerville C, Preiss J: **Isolation and Characterization of a Starchless Mutant of Arabidopsis thaliana (L.) Heynh Lacking ADPglucose Pyrophosphorylase Activity.** *Plant Physiology* 1988, **86.**

174. Yu T-S, Lue W-L, Wang S-M, Chen J: **Mutation of Arabidopsis Plastid Phosphoglucose Isomerase Affects Leaf Starch Synthesis and Floral Initiation.** *Plant Physiol* 2000, **123:**319–326.

175. Shen Y, Zhou Z, Feng S, Li J, Tan-Wilson A, Qu L-J, Wang H, Deng XW: **Phytochrome A Mediates Rapid Red Light–Induced Phosphorylation of Arabidopsis FAR-RED ELONGATED HYPOCOTYL1 in a Low Fluence** *The Plant Cell* 2009, **21:**494-506.

176. Yang SW, Jang I-C, Henriques R, Chua N-H: **FAR-RED ELONGATED HYPOCOTYL1 and FHY1-LIKE Associate with the Arabidopsis Transcription Factors LAF1 and HFR1 to Transmit Phytochrome A Signals for Inhibition of Hypocotyl Elongation.** *The Plant Cell* 2009, **21:1341-1359.**

177. Gunawan R, III FJD: **Isochron-Based Phase Response Analysis of Circadian Rhythms** *Biophysical Journal* 2006, **91:**2131-2141

178. Oprisan SA, Thirumalai V, Canavier CC: **Dynamics from a Time Series: Can We Extract the Phase Resetting Curve from a Time Series? .** *Biophysical Journal* 2003, **84:**2919-2928

179. Kunkel T, Neuhaus G, Batschauer A, Chua N-H, Schäfer E: **Functional analysis of yeast-derived phytochrome A and B phycocyanobilin adducts.** *The Plant Journal* 1996, **10(4):**625-636.

180. Fürst P, Hamer D: **Cooperative activation of a eukaryotic transcription factor: interaction between Cu(I) and yeast ACE1 protein.** *PNAS* 1989, **86:**5267–5271.

Appendix A. Model equations and additional figures for the phytochrome model systems

A.1. PHYA_FHL Model Reactions

$$Phyact\_c = (Ka\_R * RL + Ka\_FR * FRL) * \mathrm{Pr}\_GBD\_c \quad \textbf{(A.1)}$$

$$Phyact\_f = (Ka\_R * RL + Ka\_FR * FRL) * \mathrm{Pr}\_GBD\_f \quad \textbf{(A.2)}$$

$$Phyinact\_c = (Ki\_R * RL + Ki\_FR * FRL) * Pfr\_GBD\_c \quad \textbf{(A.3)}$$

$$Phyinact\_f = (Ki\_R * RL + Ki\_FR * FRL) + Pfr\_GBD\_f \quad \textbf{(A.4)}$$

$$Darkrev\_c = K\_rev * Pfr\_GBD\_c \quad \textbf{(A.5)}$$

$$Darkrev\_f = K\_rev * Pfr\_GBD\_f \quad \textbf{(A.6)}$$

$$Darkrev\_n = K\_rev * Pfr\_GBD\_n \quad \textbf{(A.7)}$$

$$\mathrm{Pr}\_GBDtrs = K3 * \mathrm{Pr}\_GBD\_f \quad \textbf{(A.8)}$$

$$Pfr\_FHLcplx = K1 * Pfr\_GBD\_f * FHL\_GAD \quad \textbf{(A.9)}$$

$$Darkrev\_cplx = K\_rev * Pfr\_FHL \quad \textbf{(A.10)}$$

$$Pfr\_FHL\_fcon = (Ki\_R * RL + Ki\_FR * FRL) * Pfr\_FHL \quad \textbf{(A.11)}$$

$$\mathrm{Pr}\_FHLdiss = K\_dis * \mathrm{Pr}\_FHL \quad \textbf{(A.12)}$$

$$luctrnbasic = Kbase \quad \textbf{(A.13)}$$

$$luctrnind = \frac{n\_Luc * (Pfr\_PIF3 + \mathrm{Pr}\_PIF3)^{a\_Luc}}{g\_Luc^{a\_Luc} + (Pfr\_PIF3 + \mathrm{Pr}\_PIF3)^{a\_Luc}} \quad \textbf{(A.14)}$$

$$LUCtrl = p\_Luc * luc \quad \textbf{(A.15)}$$

$$LuM = k\_cat * LUC * \frac{Scyt}{Km + Scyt} * RLU \quad \textbf{(A.16)}$$

$$luc\,\mathrm{deg} = m\_luc * luc \quad \textbf{(A.17)}$$

$$LUC\deg = m\_LUC * LUC \quad \textbf{(A.18)}$$

$$LUCS\deg = m\_LUC1 * LUC * Scyt \quad \textbf{(A.19)}$$

Conservation laws

$$\Pr\_GBD\_c = Pool\_Phy - Pfr\_FHL - \Pr\_FHL - Pfr\_GBD\_c - Pfr\_GBD\_f - \Pr\_GBD\_f$$

$$FHL\_GAD = Pool\_FHL - Pfr\_FHL - \Pr\_FHL$$

A.2. Final parameter set for the PhyA-FHL model

Table A.1 Final parameter set for PhyA-FHL model

Parameter description	Parameter name	Parameter value	Units	Literature data
Total PhyA concentration	Pool_PhyA	4.93E+01	nM	
Total FHL concentration	Pool_FHL	1.00E+02	nM	
Rate of assocoation for PhyA and FHL	K1	1.50E+00	$nM^{-1} * h^{-1}$	
Rate of basic (background) transcription for GAL4 promoter	k_base	4.50E-01	$nM\ h^{-1}$	
Rate transition to free pool	K2	1.00E-02	h^{-1}	
Rate of sequastration to SAPs	K3	9.90E+00	h^{-1}	
Rate of Pr_FHL complex dissociation	K_dis	3.00E-02	h^{-1}	
Rate of dark reversion	K_rev	2.00E-01	h^{-1}	0.027-1.2[76, 179]
Crossection of photoactivation by R light	Ka_R	4.96E-03	$m^2\ \mu mol^{-1}$	4.96E-03[116]
Crossection of photoactivation by FR light	Ka_FR	3.55E-05	$m^2\ \mu mol^{-1}$	3.55E-05[116]
Crossection of photoinactivation by R light	Ki_R	7.44E-04	$m^2\ \mu mol^{-1}$	7.44E-04[116]
Crossection of photoinactivation by FR light	Ki_FR	1.70E-03	$m^2\ \mu mol^{-1}$	1.70E-03[116]
Rate of luciferase RNA degradatin	m_luc	3 (12 min)	h^{-1}	
Rate of luciferase translation	p_luc	1.00E+01	h^{-1}	
Rate of luciferase protein degradation	m_LUC	0.8 (48min)	h^{-1}	0.23-1.3[119, 125]
Rate of the luciferin- dependent luciferase degradation	m_LUC1	0.00E+00	$M^{-1} * h^{-1}$	
Luciferase activity constant	k_kat	4.99E+02	$nM^{-1} * h^{-1}$	144[119]
Michaelis constant for luciferin	Km	1.90E+00	mM	0.1-2 [119]
Hill coefficient for transcription	a_luc	2.00E+00		
Maximum transcription rate	n_luc	2.94E+01	$nM\ h^{-1}$	
Light conversin factor	RLU	1.00E+01		
Diffusion coefficient	D	3.18E-06	$m^2\ h^{-1}$	
Michaelis constant for transcription	g_luc	1.04E+00	nM	

A.3. Model for PhyB_ PIF3 system (via Shimizu –Sato)

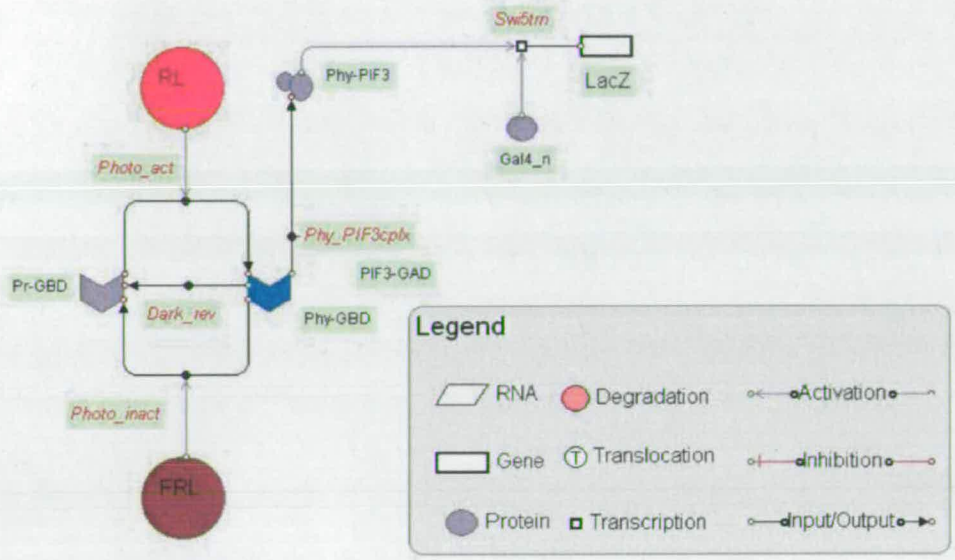


Figure A.1 Detailed structure of PhyB_ PIF3 light-switch model

A.3.1. List of reactions described in PhyB-PIF3 model

The model for the Shimizu-Sato's system (PhyB_PIF3) has the similar structure to the PhyA_FHL one, but differs in reporter –LacZ. The model lacks a description of the reporter protein kinetics due to the stability of LacZ protein and the overall relative shortness of the timescale investigated in the paper (2h).

The model reactions are presented below. Here, *Phy_GBD*, *PIF3_GAD* and *Phy_PIF3* correspond to the respective pools of active (Pfr) phytochrome B and PIF3 and their complex. *Phy_PIF3cplx*, *Phy_PIF3diss* are the reactions of association of the interacting proteins and their dissociation via photoconversion of the PhyB into inactive form. Reactions *Photo_act*, *Photo_inact* and *Dark_rev* describe the photoconversion of Pr to Pfr and vise versa and also the Pfr to Pr light independent transition. Reactions for the reporter include the transcription (*LacZ_trn*), translation (*LacZ_trl*) and degradation of the LacZ mRNA (*rLacZ_deg*)

$$Phy\_PIF3_{cplx} = K1 * [Phy\_GBD] * [PIF3\_GAD] \quad \text{(A.3.1.1)}$$

$$Phy\_PIF3_{diss} = (Ki\_R * RL + Ki\_FR * FRL) * Phy\_PIF3 \quad \text{(A.3.2.2)}$$

$$LacZ_{trn} = \frac{n\_LacZ * Phy\_PIF3^{a\_LacZ}}{g\_LacZ^{a\_LacZ} + Phy\_PIF3^{a\_LacZ}} \quad \text{(A.3.2.3)}$$

$$rLacZ_{deg} = m\_rLacZ * rLacZ \quad \text{(A.3.2.4)}$$

$$rLacZ_{trl} = p\_rLacZ * rLacZ \quad \text{(A. 3.2.5)}$$

$$Photo_{act} = (Ka\_R * RL + Ka\_FR * FRL) * \Pr\_GBD \quad \text{(A.3.2.6)}$$

$$Photo_{inact} = (Ki\_R * RL + ki\_FR * FRL) * Phy\_GBD \quad \text{(A. 3.2.7)}$$

$$Dark_{rev} = K\_rev * Phy\_GBD \quad \text{(A. 3 2 8)}$$

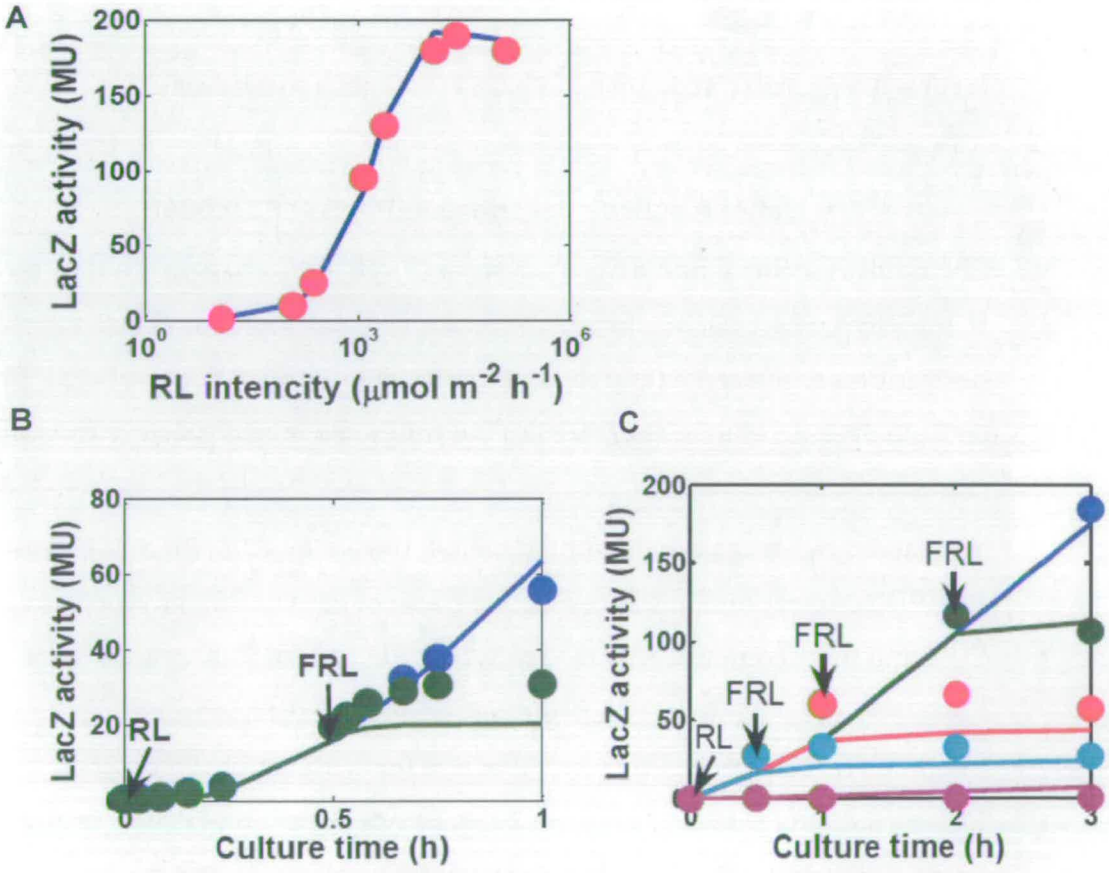


Figure A.2 Experimental data fitting results.

A Results of fitting to the dose-response curve B. Results of fitting to the peak-response curve C. Results of fitting to the activation –inhibition curve.

Table A.2 Parameters for the PhyB-PIF3 model

Parameter description	Parameter name	Parameter value	Units	Literature data
Total PhyB concentration	Pool PhyB	5.46	nM	
Total PIF3 concentration	Pool PIF3	5.46	nM	
Rate of assocoation for PhyA and FHL	K1	0.768	$nM^{-1} * h^{-1}$	[75]
Rate of dark reversion	K_rev	0.05	h^{-1}	0.027-1.2[76, 179]
Crossection of photoactivation by R light	Ka_R	4.96E-03	$^{2}\ \mu mol^{-1}$	4.96E-03[116]
Crossection of photoactivation by FR light	Ka_FR	3.55E-05	$m^{2}\ \mu mol^{-1}$	3.55E-05[116]
Crossection of photoinactivation by R light	Ki_R	7.44E-04	$m^{2}\ \mu mol^{-1}$	7.44E-04[116]
Crossection of photoinactivation by FR light	Ki_FR	1.70E-03	$m^{2}\ \mu mol^{-1}$	1.70E-03[116]
Rate of LacZ RNA degradatin	m_rLacZ	9	h^{-1}	
Rate of luciferase translation	p_LacZ	29000	h^{-1}	
Rate of luciferase protein degradation	m_LUC	0.8 (48min)	h^{-1}	0.23-1.3[119, 125]
Hill coefficient for transcription	a_LacZ	2		
Maximum transcription rate	n_LacZ	63.7	$nM\ h^{-1}$	
Michaelis constant for transcription	g_LacZ	158.4	nM	

Appendix B. Final parameter sets for the models for core oscillator

Table B.1 Parameters for the Gal-based model

Parameter description	Parameter name	Parameter value	Min	Max	Source	Units
Rate of galGal80 complex association	K1_galGal80cplx	1	0.1	10	[132]	$nM^{-1} h^{-1}$
Rate of galGal80 complex dissociation	K_galGal80diss	0.063	0.0063	0.63	[132]	nM
Rate of Gal4Gal80 complex association	K3_Gal4Gal80cplx	15.36	1.536	153.6	[132]	$nM^{-3} h^{-1}$
Rate of Gal4Gal80 complex dissociation	K_Gal4Gal80diss	0.05	0.005	0.5	[132]	nM^{3}
Rate of Gal80 export from the nucleus	K5	10	1	100		h^{-1}
Rate of Swi5 export from the nucleus	K6	10	1	100		h^{-1} nM
Distribution coefficient for Gal80	Knuc_dis	0.4	0.04	4	[132]	h^{-1}
Rate of Swi5 import to the nucleus	K_6	2	0.2	100		h^{-1}
Hill coefficient of activation by protein Swi5	a_Gal80	1.3	1	8	[132]	
Hill coefficient of activation by protein Gal4	a_Swi5	1.9	1	8		
Light response activation constant	b_Swi5	0.5	1	8		
Michaelis constant of activation by protein Swi5	g_Gal80	0.251189	0.0001	100		nM
Michaelis constant of activation by protein Gal4	g_Swi5	5.01187	0.0001	100		nM
Period of light-dark cycle	l_day	24	24	24		h
Length of the day(from dawn to dusk)	l_light	12	12	12		h
Light input(0 on dark and 1 on light)	lightresp	0	0	0		
Rate constant of degradation of Gal80p	m_Gal80	0.1	0	10		h^{-1}
Rate constant of degradation of Swi5p	m_Swi5	0.2	0	10		h^{-1}

Rate constant of rGal80 degradation	m_rGal80	2.85	0	10		h^{-1}
Rate constant of rSwi5 degradation	m_rSwi5	2.6	0	10		h^{-1}
Rate constant of r_Ssn6 degradation	m_r_Ssn6	3.85	0	10		h^{-1}
Maximum Gal80 transcription rate	n_Gal80	0.0619158	0	100		nM h^{-1}
Maximum Swi5 transcription	n_Swi5	0.0126543	0	100		nM h^{-1}
Rate constant of Gal80 translation	p_Gal80	1	0	100		h^{-1}
Rate constant of Swi5 translation	p_Swi5	1	0	100		h^{-1}
Total amount of Gal4	Pool 1	5.46	0	100	[132]	nM
Total amount of galactose	Pool 2	5000	0	100	[132]	nM

Table B.2 Parameters for the Tet-ON-based model

Parameter description	Parameter name	Parameter value	Min	Max	Source	Units
Rate of Cu_Ace1 complex association	K_Cu_cplx	1	1	100		$nM^{-4} h^{-1}$
Rate of Cu_Ace1 complex dissociation	K_Cu_diss	1	1	100		h^{-1}
Maximum rtTA transcription rate	n_rtTA	10	1	1000		$nM\ h^{-1}$
Hill coefficient of activation by protein Cu_Ace1 copmlex	a_rtTA	4	1	8	[180]	
Hill coefficiaent of inhibition by Ssn6	a_i_Ssn6	4	1	8		
Michaelis constant of inhibition	Ki	0.01	0.0001	1		nM
Rate constant for r_rtTa degradation	m_r_rtTA	4.33	0.01	10	[139]	h^{-1}
Rate constant for rtTa translation	p_rtTA	1	0	100		h^{-1}
Rate of rtTa homodimerization	Kdim_cplx	1	0	1000		$nM^{-1} h^{-1}$
Rate of rtTA homodimer dissociation	Kdim_diss	1	0	1000		h^{-1}
Rate of rtTA homodimer translocation into the nucleus	K1	1	0	1000		h^{-1}
Rate of rtTA homodimer translocation from the nucleus	K2	1	0	1000		h^{-1}
Rate of dox-rtTAdim association for the 2order reaction	Kdox_cplx	1.189	0.1189	11.89	[147-149]	$nM^{-1} h^{-1}$
Rate of dox-rtTAdim association for the 3order reaction	Kdox_cplx	1.189	0.1189	0.7	[148, 149]	$nM^{-2} h^{-1}$
Rate of dox-rtTA dim dissociation	Kdox_diss	0.07	0.007	10		h^{-1}
Rate of reaction for rtTA homodimer degaradation	m_rtTAdim	0.1	0	1000		h^{-1}
Maximum Ssn6 transcription rate	n_Ssn6	10	0	8		$nM\ h^{-1}$
Hill coefficient for activation by rtTA	a_Ssn6	2	0	1000		
Michaelis constant for transcription of Ssn6	g_Ssn6	4	0	1000		nM
Michaelis constant for transcription of rtTA	g_rtTA	1	0	1000		nM
Rate constant of r_Ssn6 degradation	m_r_Ssn6	3.85	0.1	1000		h^{-1}

Rate of Ssn6 translation	p_Ssn6	1	0	10	[133]	h^{-1}
Rate of muclear Ssn6 protein degradation	m_Ssn6_n	1	0.1	10		h^{-1}
Rate constant for Ssn6 protein translocation	K3	1	0.1	1000		h^{-1}
Ace1 total concentration	Acetot	40.359	1	1000		nM
Ace1 total concentration	doxtot	225.0047	1	100	[150, 180]	nM
Copper total concentration	Cutot	8	1	5000	[139]	nM
Rate constant for Ssn6 protein translocation from the nucleus	K4	1	0	1000	[180]	h^{-1}
Rate of cytosolic Ssn6 protein degradation	m_Ssn6_c	0.0126543	0.1			h^{-1}

Table B.3 The optimization results of the parameter set for 3order system with the basic transcription rate for rtTA

Parameter name	Units	Initial value	1st opt	2nd opt	3rd opt	4th opt	n_rtTa_ opt	a_rtTA_opt	m_r_rtTA_opt	final
K_Cu_cplx	$nM^{-4}h^{-1}$	1					1			1
K_Cu_diss	h^{-1}	1					1			1
n_rtTA	$nM\ h^{-1}$	10		10000			486			486
a_rtTA		4					6	3.07		3
g_rtTA	nM	1					0.2			0.2
Ki	nM	0.01				0.09	0.07			0.07
a_i_Ssn6		4					4.5			4.5
m_r_rtTA	h^{-1}	4.33			25		25		4.54	4.54
p_rtTA	h^{-1}	1					0.675		0.13	0.13
Kdim_cplx	$nM^{-1}h^{-1}$	1					0.03			0.02
Kdim_diss	h^{-1}	1					0.3			0.29
K1	h^{-1}	1					1			1
K2	h^{-1}	1					14.2			14.1
Kdox_cplx	$nM^{-2}h^{-1}$	1.189					0.7			0.69
Kdox_diss	h^{-1}	0.07	0.7				0.26			0.26
m_rtTAdim	h^{-1}	0.1					25			25
n_Ssn6	$nM\ h^{-1}$	10					135			135
a_Ssn6		2					4			4
g_Ssn6	nM	4					0.87			0.87
m_r_Ssn6	h^{-1}	3.85					1.44			1.44
p_Ssn6	h^{-1}	1					1.5			1.5
m_Ssn6_c	h^{-1}	0.01					0.4			0.4
m_Ssn6_n	h^{-1}	1					1			1
K3	h^{-1}	1					0.7			0.7
K4	h^{-1}	1					0.25			0.25
Acetot	nM	40.36					40.39			40.39
Doxtot	nM	225.0					1.08			1.08
Cutot	nM	8					8			8
baseTR	$nM\ h^{-1}$	--	--	30	120	300	301			301

Appendix C. The derivation of the relationship between the constants for two mechanisms for dox-binding

We assume two alternative mechanisms for binding of the dox molecule to the rtTA dimmer. The first one corresponds to simultaneous dox binding as it presented by the reaction (I);

$$A + 2B \Leftrightarrow AB_2 \qquad \text{(I)}$$

The second one describes the sequential dox binding as describes by the reactions (II) and (III):

$$A + B \Leftrightarrow AB \qquad \text{(II.1)}$$

$$AB + B \Leftrightarrow AB_2 \qquad \text{(II.2)}$$

We are looking for the situation when the kinetics of the both processes could be described in similar way and, thus, the more complicated mechanism presented by (II) could be approximated by the simplified form of (I).

For the second case we have the rate of final complex formation in a form:

$$\frac{d[AB_2]}{dt} = k_2^{II}[AB][B] - k_{-2}^{II}[AB_2],$$

in order to reduce the second system, we need to express concentration of intermediate [AB] as a function of initial concentration of components [A] and [B]. Reasonable assumption for such simplification would be that the concentration of intermediate reach steady state very fast and then does not change. That approach corresponds to the method of stationary concentrations.

Suppose that we have k_1^{II} as a rate constant for the forward reaction (II.1) and k_{-1}^{II} as a rate constant for the correspondent back reaction. Similarly, we have k_2^{II} as a rate constant for the forward reaction (II.2), and k_{-2}^{II} for the correspondent back reaction.

Now, the changing concentration of the state [AB] could be described with the following equation:

$$\frac{d[AB]}{dt} = 0 = k_1^{II}[A][B] - k_{-1}^{II}[AB] + k_{-2}^{II}[AB_2] - k_2^{II}[AB][B] \quad (1)$$

$$k_1^{II}[A][B] + k_{-2}^{II}[AB_2] = k_{-1}^{II}[AB] + k_2^{II}[AB][B]$$

$$[AB](k_{-1}^{II} + k_2^{II}[B]) = k_1^{II}[A][B] + k_{-2}^{II}[AB_2]$$

$$[AB] = \frac{k_1^{II}[A][B] + k_{-2}^{II}[AB_2]}{k_{-1}^{II} + k_2^{II}[B]} \quad (2)$$

The differential equation for the state AB_2 from the reaction (II.2) has a form:

$$\frac{d[AB_2]}{dt} = k_2^{II}[AB][B] - k_{-2}^{II}[AB_2]$$

(3)

Meanwhile, the differential equation for the same state from (I) has a form:

$$\frac{d[AB_2]}{dt} = k_1^{I}[A][B]^2 - k_{-1}^{I}[AB_2] \quad (4)$$

If we want to bring the (3) to the form of (4) we need to elimate term [AB] which is possible by substitution of (2):

$$\frac{d[AB_2]}{dt} = k_2^{II}\frac{k_1^{II}[A][B]}{k_{-1}^{II} + k_2^{II}[B]}[B] + k_2^{II}\frac{k_{-2}^{II}[AB_2]}{k_{-1}^{II} + k_2^{II}[B]}[B] - k_{-2}^{II}[AB_2] \quad (5)$$

We obtain now the first term in the form similar to the first term of (4), and we need to eliminate second term which may happen only in case when it is negligibly small compared to the third term, so that:

$$\frac{k_2^{II}k_{-2}^{II}[AB_2][B]}{k_{-1}^{II} + k_2^{II}[B]} << k_{-2}^{II}[AB_2] \quad (6)$$

$$\frac{k_2^{II}[B]}{k_{-1}^{II}+k_2^{II}[B]} << 1$$

$$\frac{k_2^{II}}{\frac{k_{-1}^{II}}{[B]}+k_2^{II}} << 1$$

$$\frac{1}{\frac{k_{-1}^{II}}{[B]k_2^{II}}+1} << 1$$

$$\frac{k_{-1}^{II}}{k_2^{II}[B]}+1 >> 1 \qquad (7)$$

We can neglect 1 as we have the expression on the left hand side much large that on the right, so we obtained:

$$k_{-1}^{II} \quad >> \quad k_2^{II}[B] \qquad (8)$$

Then we can write [AB] as

$$[AB]=\frac{k_1^{II}[A][B]}{k_{-1}^{II}}+\frac{k_{-2}^{II}[AB_2]}{k_{-1}^{II}} \qquad (9)$$

As the ratio $\frac{k_1^{II}}{k_{-1}^{II}}$ corresponds to the equilibrium constant, say, K_1^{II}, we have the expression for [AB] in a form:

$$[AB]=K_1^{II}[A][B]+\frac{k_{-2}^{II}[AB_2]}{k_{-1}^{II}} \qquad (10)$$

The differential equation (3) now takes a form of:

$$\frac{d[AB_2]}{dt}=k_2^{II}K_1^{II}[A][B]^2+\frac{k_2^{II}k_{-2}^{II}[AB_2][B]}{k_{-1}^{II}}-k_{-2}^{II}[AB_2] \qquad (11)$$

and according to (6) we can neglect second term to express final equation as:

$$\frac{d[AB_2]}{dt}=k_2^{II}K_1^{II}[A][B]^2-k_{-2}^{II}[AB_2] \qquad (12)$$

Now, (11) corresponds to (4) if

$k_1^I = k_2^{II}K_1^{II}$ and $k_{-1}^I = k_{-2}^{II}$

Appendix D. Model structure and lists of genes and reactions for the starch model

D.1. Proteins and their corresponding genes for the starch metabolic pathway

Table D.1 Proteins and their corresponding genes for the starch metabolic pathway

EC name	Definition	Gene Name	ORF[1]
4.1.2.13	Fructose-bisphosphate aldolase	fbal	Ot01g03020
		fbal	Ot03g00610
		fbalI	Ot10g01490
3.1.3.11	Fructose-1,6-bisphosphatase		Ot03g00330
			Ot14g01140
5.3.1.1	Triosephosphate isomerase		Ot09g00080
5.3.1.9	Glucose-6-phosphate isomerase		Ot11g02980
5.4.2.2	Phosphoglucomutase		Ot15g02630
2.7.7.27	ADP-glucose pyrophosphorylase	AGPLU1	Ot07g03280
		agpsu1	Ot07g02930
		agplu2	Ot20g00490
2.4.1.21	Soluble Starch synthase	SSIII-C	Ot06g03410
		SSIII-B	Ot13g01250
		SSII	Ot16g02790
		SSI	Ot13g01230
		SSIII-A	Ot16g01560
2.4.1.242	Granule-bound starch synthase	gbssI	Ot06g03200
2.4.1.18	Starch branching enzyme		Ot03g00840
		SBEI	Ot04g04110
2.4.1.1	Starch phosphorylase		Ot11g00280
			Ot04g02110
		SPho1	Ot11g01020
2.4.1.25	Disproportionating enzyme	DPE1	Ot02g05750
			Ot11g02290
			Ot11g02300
3.2.1.1	α- amylase	Aamy1	Ot16g00380
		Aamy2	Ot10g00260

[1] Genes with microarray data available are coloured

		Aamy3	Ot07g02010
3.2.1.2	β amylase		Ot02g06980
		Bamy2	Ot03g03190
		bamy1	Ot03g03170
3.2.1.68	Isoamylase (debranching enzyme)	dbe1	Ot14g02550
		dbeII	Ot12g00310
3.2.1.142	Pullalonase	spu	Ot01g03030
2.7.9.4	α-glucan, water dikinase (GWD)	SR1-A	Ot13g01510
		spr1b	
		R1 C	Ot16g02370
		spr1a	
			Ot04g04170
			Ot08g01260
	Maltose transporter	MEX1	'Ot09g03160'

D.2. The model starch metabolic pathway representation

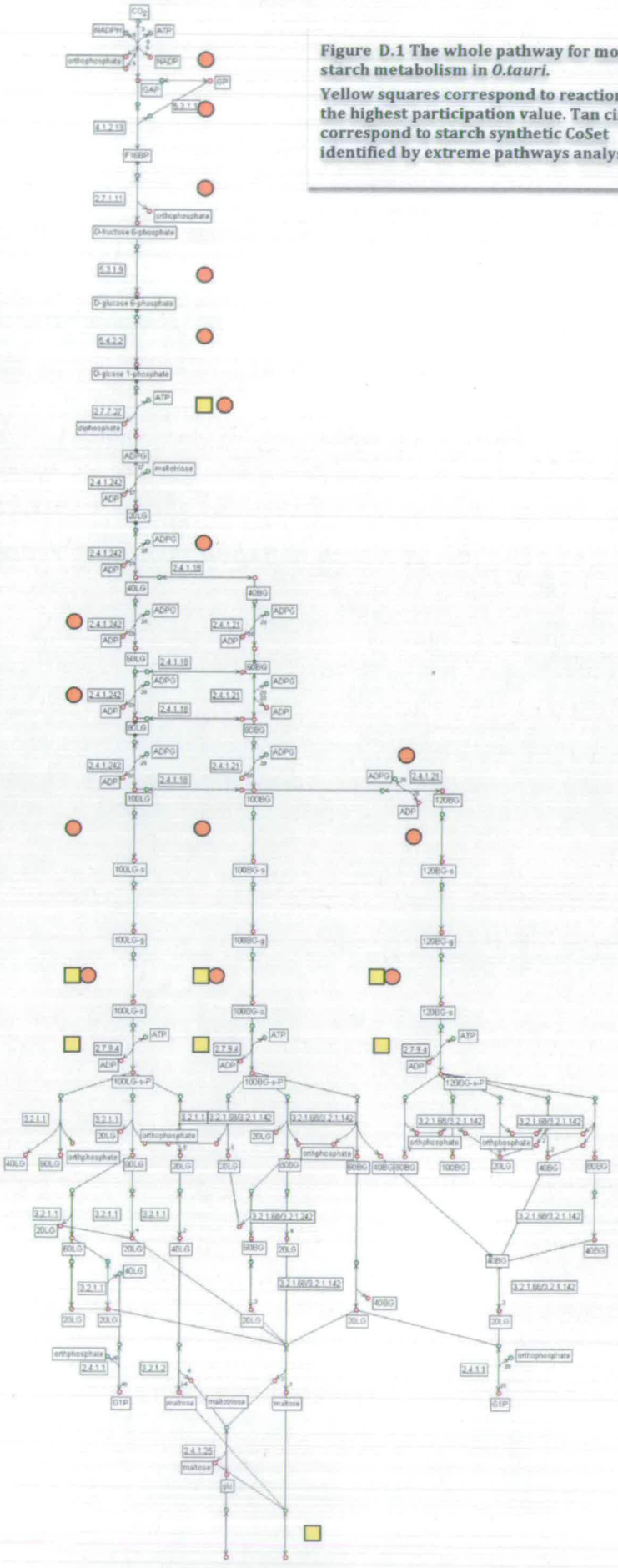


Figure D.1 The whole pathway for model starch metabolism in *O.tauri*.

Yellow squares correspond to reaction with the highest participation value. Tan circles correspond to starch synthetic CoSet identified by extreme pathways analysis

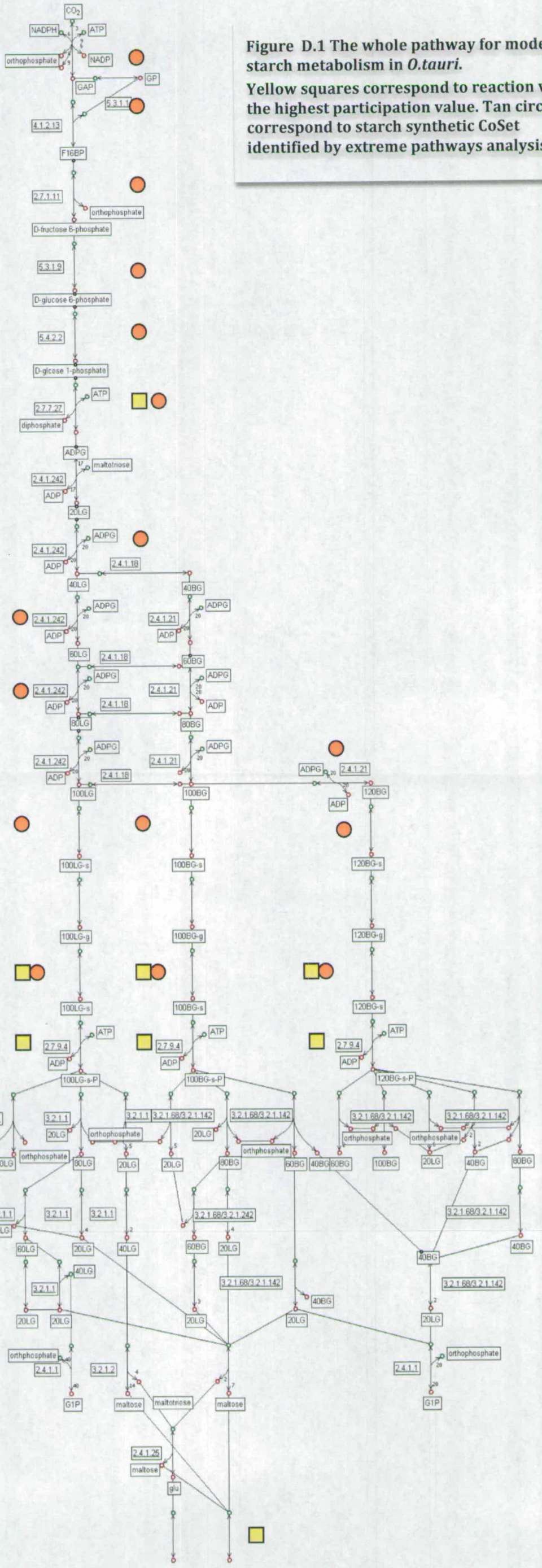


Figure D.1 The whole pathway for model starch metabolism in *O.tauri*.

Yellow squares correspond to reaction with the highest participation value. Tan circles correspond to starch synthetic CoSet identified by extreme pathways analysis

D.3. List of reaction for the starch metabolic model

Table D.2 List of reactions for the starch metabolic model with upper and low bound constraints

(* suffixes's' and 'g' correspond to 'surface' and 'granule', respectively)

reaction name	reaction	Upper flux bound, RFU e6	Low flux bound, RFU	Weight in the cost function e4	EC
'CO2'	-> CO2	0.08	0	0.0001	
'GAPex'	GAP <=>	0.0192	0	0	
'NADPHex'	NADPH <=>	1	-1000000	0	
'NADP'	NADP <=>	1	-1000000	0	
'GAPphoto'	3 CO2 + 6 NADPH + 9 ATP -> GAP + 6 NADP + 9 ADP + 8 P	1	0	0	
'GAPiso'	GAP <=> GP	1	-100000	0	5.3.1.1
'F16BP'	GAP + GP -> F16BP	0.00074	0	0	4.1.2.13
'F6P'	F16BP -> P + F6P	0.00074	0	0	3.1.3.11
'G6P'	F6P -> G6P	0.00074	0	0	5.3.1.9
'G1P'	G6P -> G1P	0.00074	0	0	5.4.2.2
'ATPex'	ATP <=>	1	-1000000	0	
'ADPout'	ADP ->	1	-1000000	0	
'Pex'	P <=>	1	-1000000	0	
'PPitoP'	PPi -> 2 P	1	-1000000	0	
'ADPG'	ATP + G1P -> PPi + ADPG	0.00074	0	0	2.7.7.27
'20LG'	17 ADPG + maltotriose -> 17 ADP + 20LG	0.000037	0	0	2.4.1.242
'40LG'	20 ADPG + 20LG -> 20 ADP + 40LG	1.85E-05	0	0	2.4.1.21
'40BGfrom40LG'	40LG -> 40BG	1.85E-05	0	0	2.4.1.18
'60BGfrom40BG'	20 ADPG + 40BG -> 20 ADP + 60BG	1.23E-05	0	0	2.4.1.21
'60LGfrom40LG'	20 ADPG + 40LG -> 20 ADP + 60LG	1.23E-05	0	0	2.4.1.242
'80LGfrom60LG'	20 ADPG + 60LG -> 20 ADP + 80LG	9.25E-06	0	0	2.4.1.242
'80BGfrom60BG'	20 ADPG + 60BG -> 20 ADP + 80BG	9.25E-06	0	0	2.4.1.21
'60BGfrom60LG'	60LG -> 60BG	1.23E-05	0	0	2.4.1.18

'80BGfrom80LG'	80LG -> 80BG	9.25E-06	0	0	2.4.1.18
'100LGfrom80LG'	20 ADPG + 80LG -> 20 ADP + 100LG	7.40E-06	0	0	2.4.1.24
'100BGfrom80BG'	20 ADPG + 80BG -> 20 ADP + 100BG	7.40E-06	0	0	2.4.1.21
'100BGfrom100LG'	100LG -> 100BG	7.40E-06	0	0	2.4.1.18
'120BGfrom100BG'	20 ADPG + 100BG -> 20 ADP + 120BG	6.17E-06	0	0	2.4.1.21
'100LG s'	100LG -> 100LG-s	7.40E-06	0	0	
'100LG g'	100LG-s <=> 100LG-g	7.40E-06	-1000000	0	
'100BG s'	100BG -> 100BG-s	7.40E-06	0	0	
'100BG g'	100BG-s <=> 100BG-g	7.40E-06	-1000000	0	
'120BG s'	120BG -> 120BG-s	6.17E-06	0	0	
'120BG g'	120BG-s <=> 120BG-g	6.17E-06	-1000000	0	
'starch ex'	100LG-g + 4 100BG-g + 2 120BG-g <=>	1.12E-06	-1000000	1.046814	
'100LG s P'	P + 100LG-s -> 100LG-s-P	7.40E-06	0	0	2.7.9.4
'100BG s P'	P + 100BG-s -> 100BG-s-P	7.40E-06	0	0	2.7.9.4
'120BG s P'	P + 120BG-s -> 120BG-s-P	6.17E-06	0	0	2.7.9.4
'100LG s Pto40LG60LG'	100LG-s-P -> P + 40LG + 60LG	7.40E-06	0	0	3.2.1.1
'100LG s Pto80LG20LG'	100LG-s-P -> P + 20LG + 80LG	7.40E-06	0	0	3.2.1.1
'100LG s Pto20LG'	100LG-s-P -> P + 5 20LG	7.40E-06	0	0	3.2.1.1
'100BG s Pto20LG'	100BG-s-P -> P + 5 20LG	7.40E-06	0	0	3.2.1.68
'100BG s Pto80BG20LG'	100BG-s-P -> P + 20LG + 80BG	7.40E-06	0	0	3.2.1.14
'100BG s Pto40BG60BG'	100BG-s-P -> P + 40BG + 60BG	7.40E-06	0	0	3.2.1.68
'120BG s Pto60BG'	120BG-s-P -> P + 2 60BG	6.17E-06	0	0	3.2.1.68
'120BG s Pto20LG100BG s	120BG-s-P -> 20LG + 100BG-s-P	6.17E-06	0	0	3.2.1.68

P'					
'120BG s Pto20LG'	120BG-s-P -> P + 6 20LG	6.17E-06	0	0	3.2.1.68
'120BG s Pto20LG40BG'	120BG-s-P -> P + 2 20LG + 2 40BG	6.17E-06	0	0	3.2.1.142
'120BG s Pto80BG40BG'	120BG-s-P -> P + 40BG + 80BG	6.17E-06	0	0	3.2.1.142
'80BGto20LG'	80BG -> 4 20LG	9.25E-06	0	0	3.2.1.68
'80BGto40BG'	80BG -> 2 40BG	9.25E-06	0	0	3.2.1.142
'80LGto20LG'	80LG -> 4 20LG	9.25E-06	0	0	3.2.1.1
'80LGto40LG'	80LG -> 2 40LG	9.25E-06	0	0	3.2.1.1
'80BGto60BG20LG'	80BG -> 20LG + 60BG	9.25E-06	0	0	3.2.1.68
'80LGto60LG20LG'	80LG -> 20LG + 60LG	9.25E-06	0	0	3.2.1.1
'60BGto20LG'	60BG -> 3 20LG	1.23E-05	0	0	3.2.1.68
'60BGto40B20LG'	60BG -> 20LG + 40BG	1.23E-05	0	0	3.2.1.68
'60LGto20LG'	60LG -> 3 20LG	1.23E-05	0	0	3.2.1.1
'60LGto40LG20LG'	60LG -> 20LG + 40LG	1.23E-05	0	0	3.2.1.1
'40BGto20LG'	40BG -> 2 20LG	1.85E-05	0	0	3.2.1.68
'40LGtomaltose'	40LG -> 4 maltotriose + 14 maltose	1.85E-05	0	0	3.2.1.2
'20LGtomaltose'	20LG -> 2 maltotriose + 7 maltose	0.000037	0	0	3.2.1.2
'20LGtoG1P'	20 P + 20LG -> 20 G1P	0.000037	0	0	2.4.1.1
'40LGto G1P'	20 P + 40LG -> 20 G1P + 20LG	1.85E-05	0	0	2.4.1.1
'maltotriose disp'	20LG + 20 maltotriose -> 40LG + 10 glu	0.000247	0	0	2.4.1.25
'glucosidase'	maltotriose -> maltose + glu	0.000247	0	0	3.2.1.20
'maltose out'	maltose ->	0.000167	0	0.096333	MEX1
'glu out'	glu ->	0.000122	0	0.065003	
'20LGfromADPG'	20ADPG -> 20 ADP + 20LG	0.00037	0	0	2.4.1.21

Appendix E. Results of the Flux Variability and Robustness analysis

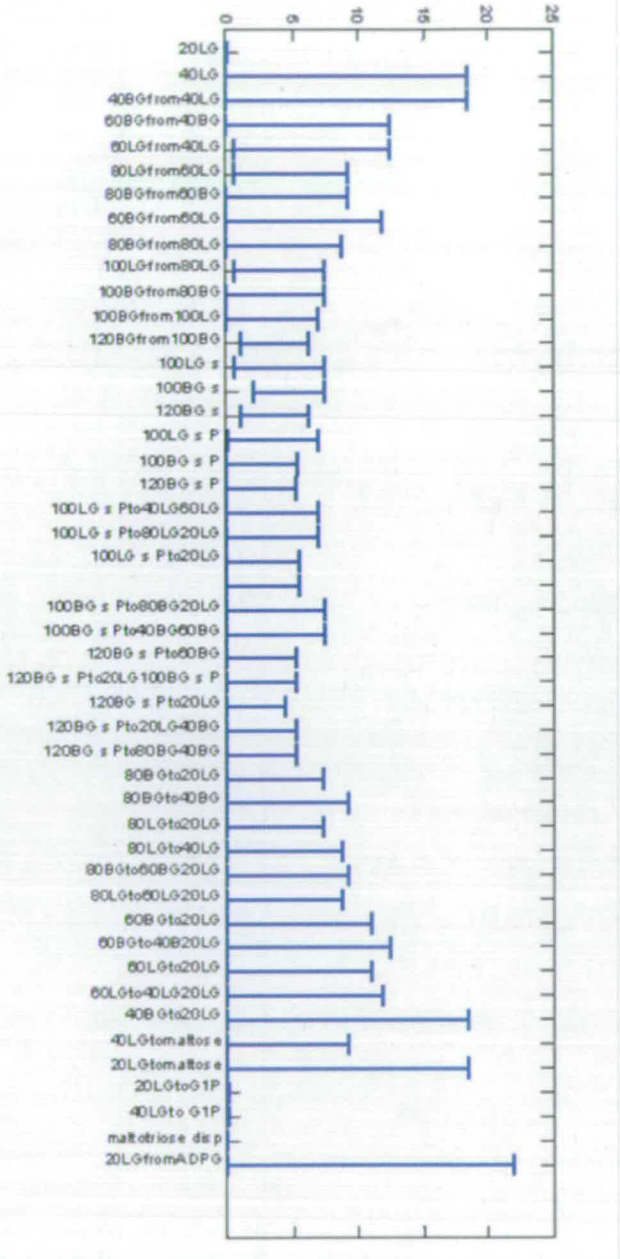
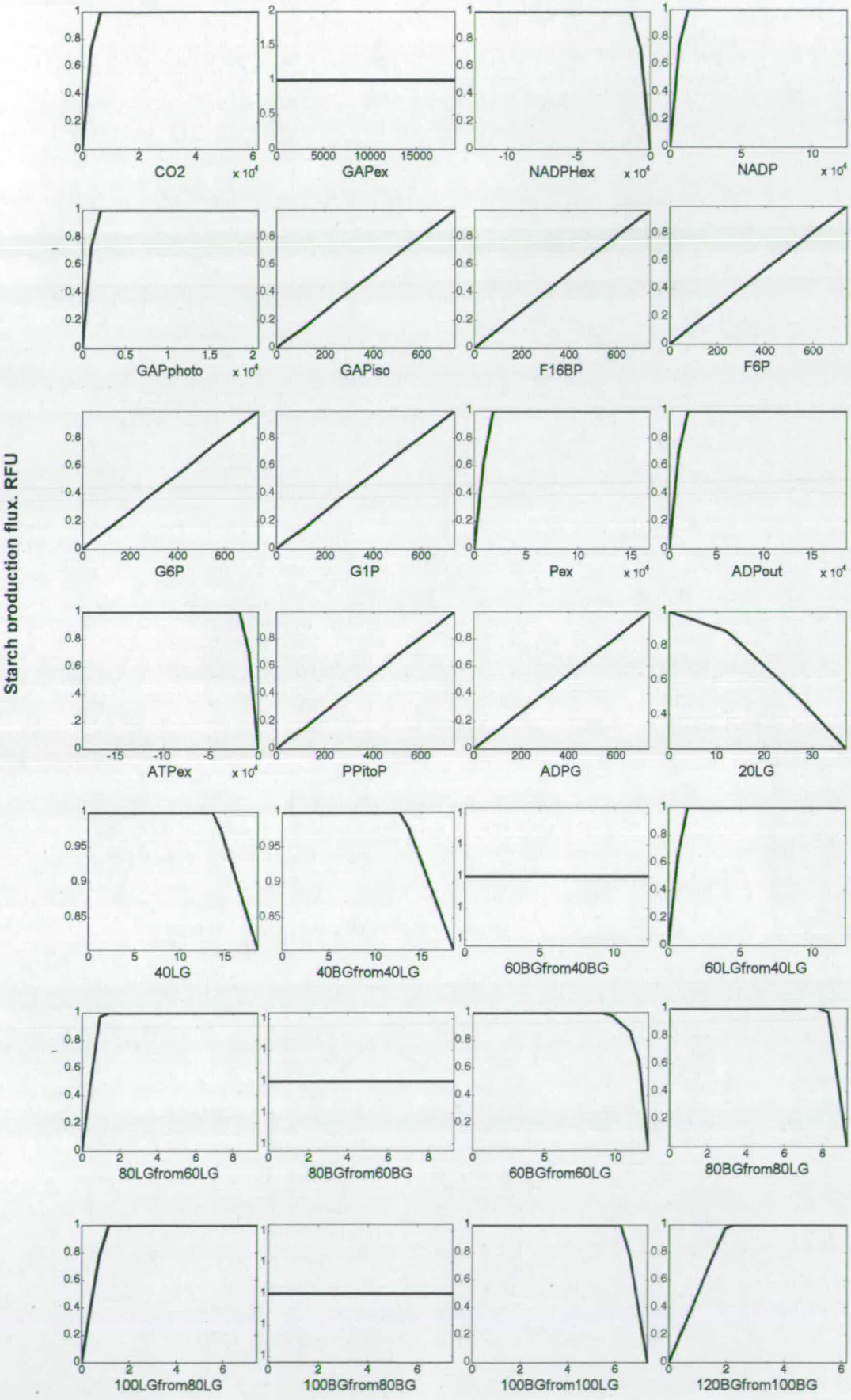
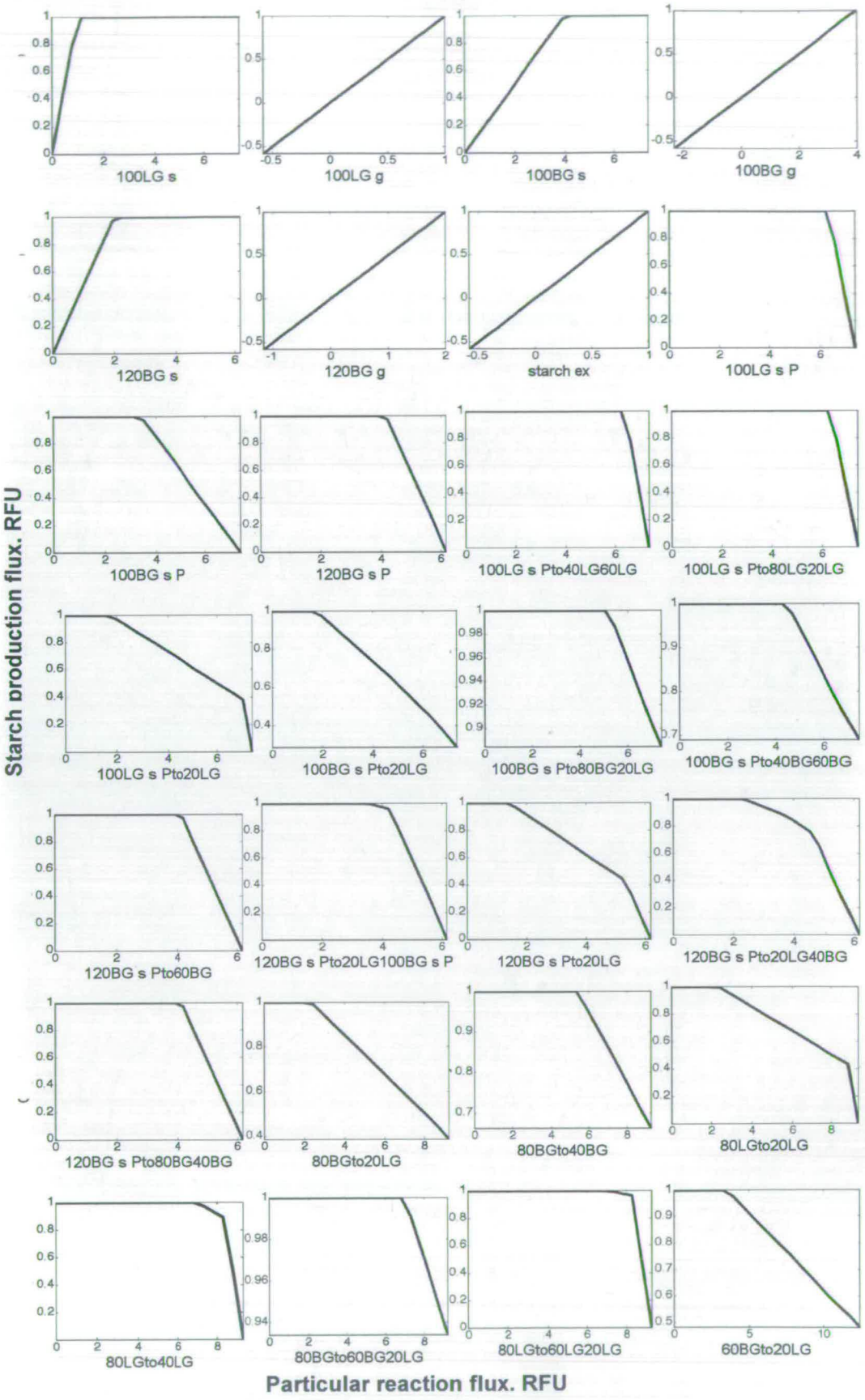


Figure E.1 Results of the FVA analysis for the model starch metabolic system. The maximum value of the objective function is computed and then used for calculation of minimum and maximum allowable flux values through each reaction.


Starch production flux, RFU
CO2
GAPex
NADPHex
NADP
GAPphoto
GAPiso
F16BP
F6P
G6P
G1P
Pex
ADPout
ATPex
PPitoP
ADPG
20LG
40LG
40BGfrom40LG
60BGfrom40BG
60LGfrom40LG
80LGfrom60LG
80BGfrom60BG
60BGfrom60LG
80BGfrom80LG
100LGfrom80LG
100BGfrom80BG
100BGfrom100LG
120BGfrom100BG
Particular reaction flux, RFU


100LG s
100LG g
100BG s
100BG g
120BG s
120BG g
starch ex
100LG s P
100BG s P
120BG s P
100LG s Pto40LG60LG
100LG s Pto80LG20LG
100LG s Pto20LG
100BG s Pto20LG
100BG s Pto80BG20LG
100BG s Pto40BG60BG
120BG s Pto60BG
120BG s Pto20LG100BG s P
120BG s Pto20LG
120BG s Pto20LG40BG
120BG s Pto80BG40BG
80BGto20LG
80BGto40BG
80LGto20LG
80LGto40LG
80BGto60BG20LG
80LGto60LG20LG
60BGto20LG
Starch production flux. RFU
Particular reaction flux. RFU

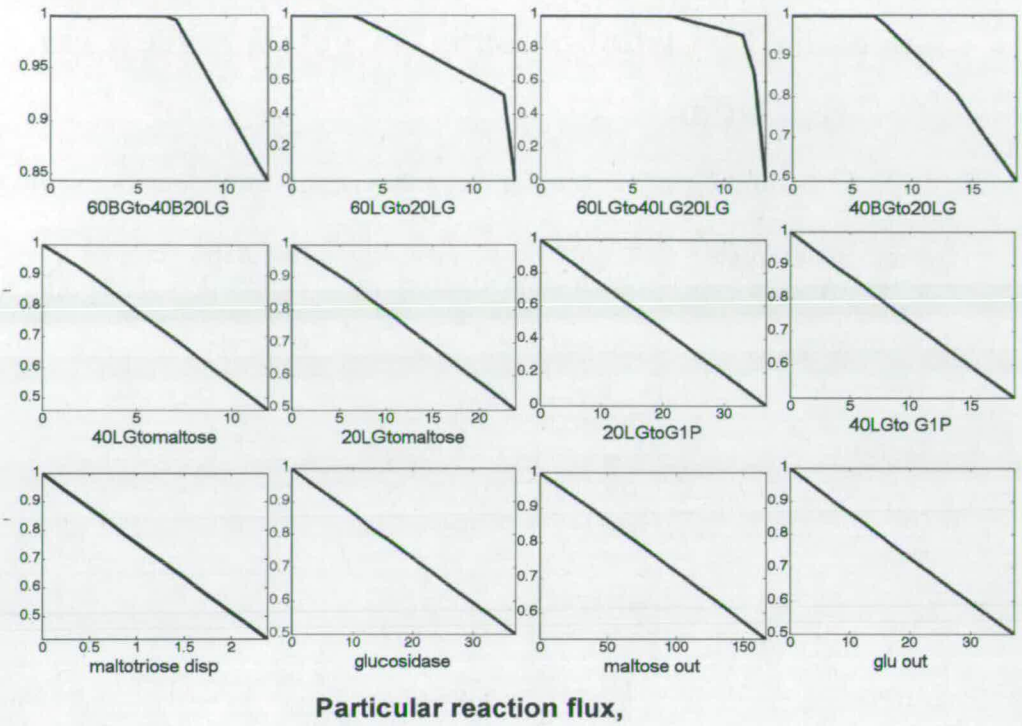


Figure E.2. Results of robustness analysis

Appendix F. List of potential targets for circadian regulation

Table F.1 Summary for targets for circadian regulation identified from different analysis

Regulation targets obtained in the robustness analysis (reactions)	Regulation targets obtained in single gene deletion analysis (genes)	Regulation targets obtained in gene overexpression analysis	One array substitution
GAPiso			
F16BP	Ot10g01490 (fructose-bisphosphate aldolase)		
F6P	Ot03g00330 (fructose-1,6-bisphosphotase)		
G6P			
G1P	Ot15g02630 (phosphoglucomutase)		
ADPG	Ot7g02930 (AGPase)		Ot07g02930 (AGPase)
20LG	Ot06g03200 (GBSSI)	Ot06g03200 (GBSSI)	
100LGs_Pto20LG			Ot07g02010 (α-amylase)
100BGs_Pto 20LG			Ot12g00310 and Ot14g02550 (isoamylases)
120BGs_Pto20LG			Ot12g00310 and Ot14g02550 (isoamylases)
80BGto20LG			Ot12g00310 and Ot14g02550 (isoamylases)
80LGto20LG			Ot07g02010 (α-amylase)
60BGto20LG			Ot12g00310 and Ot14g02550 (isoamylases)
20LGto maltose	Ot03g03190 (β–amylase)		Ot03g03190 (β-amylase)
40LFto maltose	Ot03g03190 (β–amylase)		Ot03g03190 (β–amylase)
20LGtoG1P	Ot04g02110 (starch phosphorylase)	Ot11g00280 and Ot11g01020 (phophorylase)	

40LGtoG1P	Ot04g02110 (starch phosphorylase)	Ot11g00280 and Ot11g01020 (phosphrylase)	
maltotriose_disp			
glucosidase			
maltose out	Ot09g03160 (MEX1)		
glu out		Ot14g01870 and Ot03g05590 (glucose transprtes)	
	Ot13g01510 (glucan, waterdikinase)	Ot08g01260 (glucan, water dikinase)	Ot03g01510 (glucan water dikinase)